KOÇ UNIVERSITY-TÜSİAD ECONOMIC RESEARCH FORUM
WORKING PAPER SERIES

ESTIMATION OF BINARY CHOICE MODELS WITH LINEAR INDEX AND DUMMY ENDOGENOUS VARIABLES

Nese Yıldız

Working Paper 1202
January 2012

KOÇ UNIVERSITY-TÜSİAD ECONOMIC RESEARCH FORUM
Rumelifeneri Yolu 34450 Sarıyer/Istanbul

Estimation of Binary Choice Models with Linear Index and Dummy Endogenous Variables*

Neşe Yıldız†
Department of Economics, University of Rochester.

First version: September, 2004. This version: December 18, 2011.

Abstract

This paper presents computationally simple estimators for the index coefficients in a binary choice model with a binary endogenous regressor without relying on distributional assumptions or on large support conditions and yields root-n consistent and asymptotically normal estimators. We develop a multi-step method for estimating the parameters in a triangular, linear index, threshold-crossing model with two equations. Such an econometric model might be used in testing for moral hazard while allowing for asymmetric information in insurance markets. In outlining this new estimation method two contributions are made. The first one is proposing a novel "matching" estimator for the coefficient on the binary endogenous variable in the outcome equation. Second, in order to establish the asymptotic properties of the proposed estimators for the coefficients of the exogenous regressors in the outcome equation, the results of Powell, Stock and Stoker (1989) are extended to cover the case where the average derivative estimation requires a first step semi-parametric procedure.

1 Introduction:

The estimation of econometric models with binary outcomes and a binary endogenous regressor is of considerable practical importance. Given the importance of such models, this paper focuses on the linear index threshold crossing model and presents an easy-to-implement estimation method that does not require knowledge of the parametric form of the distribution of the unobservables. In addition, the proposed method does not rest on the existence of any regressors with unbounded support.

A popular method for dealing with an endogenous regressor when the outcome is continuous and the unobservable variables enter the outcome equation additively involves estimating the parameters of interest by ordinary instrumental regression analysis. But when the outcome is binary the unobserved variables in the outcome equation cannot be additive, except within the context of the linear probability model. Consequently, as is well known, we cannot rely on the standard instrumental variables methods to get consistent estimates. Heckman (1978) and Amemiya (1978) propose a parametric solution to the problem by specifying a joint distribution, typically joint normality, of the unobserved terms and obtains estimates for the parameters of interest using maximum likelihood estimation. While this method delivers consistent estimates when the joint distribution of the error terms is correctly specified, the consistency of these estimators cannot be guaranteed when the error distribution is misspecified.[1]

*I would like to thank Takeshi Amemiya, Aprajit Mahajan, Edward Vytlacil, James Powell, and Todd Elder for very helpful comments and questions on this paper.

†Correspondence: University of Rochester, Department of Economics, Harkness Hall, Rochester, NY 14627; Email: nyildiz@mail.rochester.edu; Phone: 585-275-5782.

[1]Bhattacharya, McCaffrey and Goldman (2004) presents Monte Carlo results suggesting that bivariate probit estimates of the average effect of the binary endogenous variable on the outcome could be severely biased when the joint distribution of the unobservables is not normal.

Vytlacil and Yıldız (2007) shows that it is possible to identify and consistently estimate the average effect of a binary endogenous variable on the outcome without imposing large support conditions and without relying on parametric distributional or functional form assumptions. While the results developed there hold for a general class of models, the paper primarily focuses on one parameter, the average treatment effect.

The average treatment effect is no doubt an important parameter of interest. In some contexts, however, estimation of the joint distribution of the unobservables in outcome equation and those in the first stage equation might also be of interest, as in analyzing insurance markets, which will be discussed in Section 3 when the econometric model has been introduced. While the model in Vytlacil & Yıldız (2007) is general, the joint distribution of unobservables cannot be estimated under such a general model. This paper focuses on the case where both the outcome and endogenous regressor equation have linear index threshold crossing form, but the distribution of the unobservables is left unspecified other than some smoothness restrictions which are imposed for estimation. The paper presents a method of estimating the coefficients on all the regressors as well as the joint distribution of the unobservables.

In a broad sense, our estimator for the coefficient of the binary endogenous regressor is akin to a matching estimator. Matching is based on the idea of contrasting the outcomes of individuals for whom the binary regressor is 1 with the outcomes of comparable individuals for whom the binary regressor is 0 where comparisons are constructed on the basis of observed characteristics of individuals. However, when the binary regressor is endogenous, the basic assumption underlying matching may be untenable. Our three stage estimator suggests an alternative method for comparing individuals across the two groups. In a sense, we propose matching individuals using a different metric where the proposed metric takes the endogeneity of the binary regressor into account.

2 Literature Review:

Lewbel (2000) proposes an estimation method which does not require that the distribution of the unobservables is known and in which the endogenous variable could be either binary or continuous. This method relies on the existence of a continuously distributed special regressor with large support that influences the outcome variable (but not the endogenous regressor).[2] Despite its simplicity, this estimation procedure has the drawback of placing higher weights on observations that are from parts of the population with low probability of being observed.

Chen and Vytlacil (2005) studies a nonlinear panel data models with lagged dependent variables. The strategy for identifying the coefficient of the lagged dependent variable in their paper and the coefficient of the binary endogenous regressor in this paper are both straightforward extensions of the identification strategy in Vytlacil and Yıldız (2007). Both the motivation and estimation strategies of the two papers, however, are different. The differences in the estimation method will be explained after the discussion of the estimation method in Section 5.

An alternative approach to dealing with endogeneity, particularly when the unobservables in the outcome equation are not additively separable from the regressors, is the control function approach.

[2]This feature of the estimation method of Lewbel (2000) is related to the estimation method proposed in this paper. In particular, in our context we need some exogenous variation in the outcome equation even after conditioning on the exogenous components in the equation for the binary endogenous variable, but here none of the regressors is required to have large support. The precise nature of this relation will be made clear when the model and the identification assumptions are introduced.

Blundell and Powell (2004) uses this approach for semiparametric estimation in single index binary response models with a continuous endogenous regressor. The control function approach, however, is not applicable when the endogenous regressor is binary. Nevertheless, our estimation method bears some similarity to their method. Under their exclusion restrictions, the outcome variable can be characterized by a "multiple index regression" model, with conditional distribution of the outcome, given the regressors and the error term from the equation for the endogenous regressors, depending on the regressors only through a single index. In estimating the coefficients on the regressors in the outcome equation, they exploit the invertibility of this distribution function with respect to its first argument. In estimating the coefficient on the binary endogenous regressor, we exploit the invertibility of the same conditional distribution; however, a crucial difference arises in recovering this conditional distribution. Our analysis requires stronger assumptions than their analysis, but this is because recovering this conditional distribution is harder in the current context of a binary endogenous regressor.

Imbens and Newey (2009) use the control function approach to develop identification results the average and quantile effects of a continuous endogenous regressor in triangular simultaneous equations models.

This paper is also related to the extensive literature that considers endogenous regressors in semiparametric or nonparametric models without additive separability. Altonji and Matzkin (2005) presents methods for identifying and estimating the same effect holding the conditional distribution of the error term conditional on the covariates fixed within the context of a panel data model with nonseparable error terms and endogenous regressors. One of their methods requires that the outcome variable is strictly monotonic in the error term, while their other method assumes that the unobservable components and the covariates that determine the outcome are independent conditional on the instruments, so that conditional on the instruments there is no problem of endogeneity. Thus, neither of their estimation methods is suitable for the model of interest in this paper. Chernozhukov and Hansen (2005) study the identification of quantile treatment effects in the presence of endogeneity. Their analysis assumes that the outcome of interest is strictly increasing in the unobserved component, rendering it inapplicable to the binary outcome case. Chesher (2003) presents a method for the local identification of derivatives and partial differences of structural quantile functions in the context of a nonseparable model where the endogenous regressor is continuous.[3] Chesher (2005) studies the same problem with a discrete endogenous variable. But his results do not extend to the case in which the endogenous regressor takes only two values. In addition, Blundell and Powell (2003) provides an excellent summary of nonparametric or semiparametric estimation methods for regression models with continuous endogenous regressors.

3 The Model:

This paper examines the linear index, threshold crossing model with two equations:

$$D = 1\{Z'\gamma - U \geq 0\} \tag{1}$$

$$Y = 1\{\alpha_Y + X'\beta + D\delta - \varepsilon \geq 0\}, \tag{2}$$

where D denotes the binary endogenous variable, Y represents the outcome variable, $X = (W_1', W_2')' \in \mathbb{R}^{1+d_1+d_2}$, and $Z = (W_2', W_3')' \in \mathbb{R}^{1+d_2+d_3}$ are observed vectors of random variables, (ε, U) is an un-

[3] Ma and Koenker (2009) present estimation methods for the effects identified in Chesher (2003). Jun (2009) is a semiparametric version of Ma and Koenker (2009).

observed random vector, and (β, δ, γ) are the parameters of interest. Note that X and Z may have common components, but the identification method presented below will require existence of a continuous component of W_3, with corresponding γ coefficient not equal 0. Thus, $d_3 \geq 1$. Without loss of generality we assume that W_{3d_3} is continuous, and we normalize the absolute value of γ_{3d_3} to be 1. Similarly, identification of β will require a continuous component of either W_1 or W_2, with a non-zero β coefficient. Let β_{k^*} denote this coefficient. For these restrictions to hold $d = d_1 + d_2 + d_3$ is assumed to be greater than or equal to 2. The scale normalization we adopt for the outcome equation is that the absolute value of β_{k^*} is set equal to 1. The location normalizations we adopt are $E(U) = E(\varepsilon) = 0$. We should note, however, that identification of γ, β and δ, as well as testing whether ε and U are independent of each other do not depend on the location normalizations imposed, that is, we can do these things without knowing what α_D and α_Y are. For completeness, we discuss how α_D and α_Y can be identified under symmetry assumption on F_U and F_ε, respectively, at the end of identification section. $1\{\}$ denotes the indicator function. Moreover, for $v : \mathbb{R}^k \to \mathbb{R}$, $\frac{\partial v(s)}{\partial s'}$ denotes the $1 \times k$ dimensional gradient vector of the function v. This model has a form similar to a multivariate probit model, and is referred to as a "multivariate probit model with structural shift" by Heckman (1978).

The econometric model given by equations (1) and (2) may be useful in studying insurance markets. For example suppose we would like to estimate the probability of an accident during a policy period. In this situation, the outcome of interest is a binary variable which takes value one if the driver is involved in an accident during an insurance policy period, and the binary endogenous variable is the indicator for whether the driver has purchased comprehensive insurance or not. In such a model if the coefficient on the binary endogenous regressor is positive then we might suspect that moral hazard is an issue, as in this case the probability that the driver has an accident is higher after controlling for covariates if the driver has bought comprehensive insurance. The methods of this paper would allow the researcher to estimate this coefficient and the coefficients on exogenous regressors without requiring ε and U to be independent. If the driver knows his type, which neither the insurance company nor the econometrician know, and if this type makes the driver more accident prone and knowing this the driver is more likely to buy high coverage, then we would have adverse selection and U and ε would not be independent. Since the estimators do not require U and ε to be independent. once we have estimators for the coefficients we can also estimate the joint distribution of the unobservables and test whether U and ε are independent.

The following assumptions will be imposed throughout the paper:

(A-0) $\{Y_i, D_i, W'_{1i}, W'_{2i}, W'_{3i}\}$ $i = 1, ..., n$ is an i.i.d. sequence of random vectors.

(A-1) The distribution of (ε, U) is absolutely continuous with respect to Lebesgue measure with positive density on $\mathbb{R}^2$;

(A-2) (U, ε) is independent of (W'_1, W'_2, W'_3);

(A-3) Z has density with respect to Lebesgue measure on $\mathbb{R}^{d_2+d_3}$, $(W'_1, W'_2, W'_2\gamma_2 + W'_3\gamma_3)'$ has density with respect to Lebesgue measure on $\mathbb{R}^{d_1+d_2+1}$;

(A-4) $\text{Supp}(Z'\gamma, X'\beta) \cap \text{Supp}(Z'\gamma, X'\beta + \delta) \neq \emptyset$.

Even though assumption (A-0) is not needed for identification, it will be used for all of our estimation results. Assumptions (A-1)-(A-4) are needed for identification. Additional assumptions will be imposed in subsequent sections as they are needed for estimation. Assumption (A-1) is a regularity condition sufficient to guarantee that the relevant conditional expectations are smooth functions. Assumption (A-2) is critical for both identification and estimation. For identification, a weaker version of assumption (A-3), namely that $X'\beta$ is continuous, as opposed to all of X, would suffice; the stronger version is imposed for estimation. More specifically, this assumption implies that the exogenous regressors are continuous.[4] The second part of assumption (A-3) implies that there is an element in W_3 with a non-zero coefficient and is continuous. This part of the assumption will be used in estimating β. In addition, the second part of this assumption implies that $(X'\beta, Z'\gamma)$ has density with respect to Lebesgue measure on $\mathbb{R}^2$ as long as there is at least one β that is different from 0. Our strategy for identifying δ depends on our ability to find shifts in $X'\beta$ that exactly offset the effect of a change in D from 0 to 1. To guarantee that pairs of $X'\beta$ that "undo" the effect of a shift in D exist we need the support condition stated in assumption (A-4).[5] In addition to guaranteeing that such $X'\beta$ pairs exist, we need a way to identify such pairs. Thus, in estimating δ, it will be essential to be able to vary $Z'\gamma$ with positive probability while holding $X'\beta$ constant. This requires that there is some variation in $X'\beta$ that is not perfectly correlated with $Z'\gamma$,[6] and the second part of assumption (A-3) guarantees this as well.

4 Identification Analysis:

4.1 Identification of γ and β:

Identification of binary choice models of linear index threshold crossing form is well known. (See Manski (1988) for example.) Here we include a brief discussion of how γ is identified for completeness. By (A-3), Z is a continuous random vector, and $\frac{\partial E[D|Z=z]}{\partial z_j} = f_U(\alpha_D + z'\gamma)\gamma_j$ for each $j = 1, ..., d_2 + d_3$. Thus, the signs and the ratios of γ_j are identified, since W_{3,d_3} is assumed to be a continuous random variable with corresponding $|\gamma_{3,d_3}| = 1$. If there are discrete regressors, under additional support restrictions the coefficients of discrete regressors could be identified. To see how this can be done, suppose Z_1 is continuous and $\gamma_1 \neq 0$. Suppose there is only one other Z, Z_2, which is discrete and takes values $z_{21}, .., z_{2M}$. Then for each m, $\frac{\partial E(D|Z_1=z_1, Z_2=z_{2m})}{\partial z_1} = f_U(\alpha_D + z_1\gamma_1 + z_{2m}\gamma_2)\gamma_1$. We identify the sign of γ_1 from this, and can normalize γ_1 to be 1 if it is positive, or -1 if it is negative. Suppose γ_1 is positive. Note that $E(D|Z_1 = z_1, Z_2 = z_{2m}) = E(D|Z_1 = \tilde{z}_1, Z_2 = z_{2k})$ if and only if $z_1 + z_{2m}\gamma_2 = \tilde{z}_1 + z_{2k}\gamma_2$. Therefore, $\gamma_2 = \frac{z_1 - \tilde{z}_1}{z_{2k} - z_{2m}}$. Let $r_z := z_{2m^*} - z_{2k^*}$ where m^* and k^* attain $\min\{|z_{2m} - z_{2k}| : m \in \{1, ..., M\}, k \in \{1, ..., M\} \setminus \{m\}\}$. Then if $P(\text{Supp}(Z_1) \cap \text{Supp}(Z_1 + r_z\gamma_2)) > 0$, γ_2 is also identified.

By (A-3) again, there is an element of X that is continuous and has a non-zero β coefficient. Then

$$\frac{\partial E[Y|X = x, Z = z]}{\partial x} = \left[\frac{\partial P(U \leq \alpha_D + z'\gamma, \varepsilon \leq \alpha_Y + x'\beta + \delta)}{\partial(x'\beta)} + \frac{\partial P(U > \alpha_D + z'\gamma, \varepsilon \leq \alpha_Y + x'\beta)}{\partial(x'\beta)}\right]\beta.$$

Since the expression in square brackets is strictly positive by assumptions (A-1) and (A-2), we identify the signs and ratios of coefficients of continuous X. In this paper, we assume all the X's are continuous,

[4] Here, the estimation in the first two stages are done either using the methods developed by Powell, Stock and Stoker (1989) directly or an extended version of them. This method requires that all the regressors are continuous. Using the methods outlined in Härdle and Horowitz (1996), we can extend this analysis to the case where some of the regressors are discrete. For this extension only one of the regressors must be continuous.

[5] Assumption (A-4) may look strange given that the coefficients are only identifiable up to scale; however if the assumption holds for some $\big(\gamma, (\beta, \delta)\big)$ it will also hold for $\big(a_1\gamma, a_2(\beta, \delta)\big)$ for any pair of non zero constants (a_1, a_2).

[6] This feature of our estimation method is similar to Lewbel (2000).

but if there are discrete regressors and if the supports of the continuous X's are rich enough, we can identify the coefficients on the discrete regressors in a way that is similar to the example given in the discussion of identification of γ. The problem with the above equation is that it assumes we can vary each component of X while holding Z constant. If X and Z have common components, we cannot do that. Given that γ is identified, we could repeat the same argument by considering $E[Y|X = x, Z'\gamma = t]$ instead.[7] This would give us:

$$\frac{\partial E[Y|Z'\gamma = t, X = x]}{\partial x} = \left[\frac{\partial P(U \leq \alpha_D + t, \varepsilon \leq \alpha_Y + x'\beta + \delta)}{\partial(x'\beta)} + \frac{\partial P(U > \alpha_D + t, \varepsilon \leq \alpha_Y + x'\beta)}{\partial(x'\beta)}\right]\beta. \tag{3}$$

Under the second part of assumption (A-3), equation (3) is valid, and using this equation we can identify signs and scales of components of β.

4.2 Identification of δ:

The strategy that we will use to identify δ is the same as in Vytlacil and Yıldız (2004,2007) and Chen and Vytlacil (2005). This strategy is based on finding shifts in $X'\beta$ which directly compensate for a shift in D to identify δ. Given our model and assumptions, we can use variations in $Z'\gamma$ to identify such $X'\beta$ shifts.

Based on results presented in the previous subsection, in studying identification of δ, we are going to assume that γ and β are known. If D and X were independent of ε, we could identify δ using arguments similar to those given when identification of coefficients of discrete Z was discussed in the previous section. In our problem W (and hence X) is independent of ε, but D is not. Even though we can compute $E(Y = 1|X'\beta = x'\beta, D = 1)$ and $E(Y = 1|X'\beta = \tilde{x}'\beta, D = 0)$, when D is endogenous these no longer give us the probabilities, $P(\varepsilon \leq \alpha_Y + x'\beta + \delta)$ and $P(\varepsilon \leq \alpha_Y + \tilde{x}'\beta)$, that we need to identify δ. Nevertheless, we can use exogenous variation in $Z'\gamma$ to identify different $X'\beta$ values that compensate for the effect of a change in D from 0 to 1. To see how this can be done, note that since Y, D, X and Z are observed, and since γ and β are known, using the data we can compute

$$E[DY|X'\beta = \overline{x}'\beta, Z'\gamma = z'\gamma] = P(U \leq \alpha_D + z'\gamma, \varepsilon \leq \alpha_Y + \overline{x}'\beta + \delta)$$

and

$$E[(1-D)Y|X'\beta = \overline{x}'\beta, Z'\gamma = z'\gamma] = P(U > \alpha_D + z'\gamma, \varepsilon \leq \alpha_Y + \overline{x}'\beta).$$

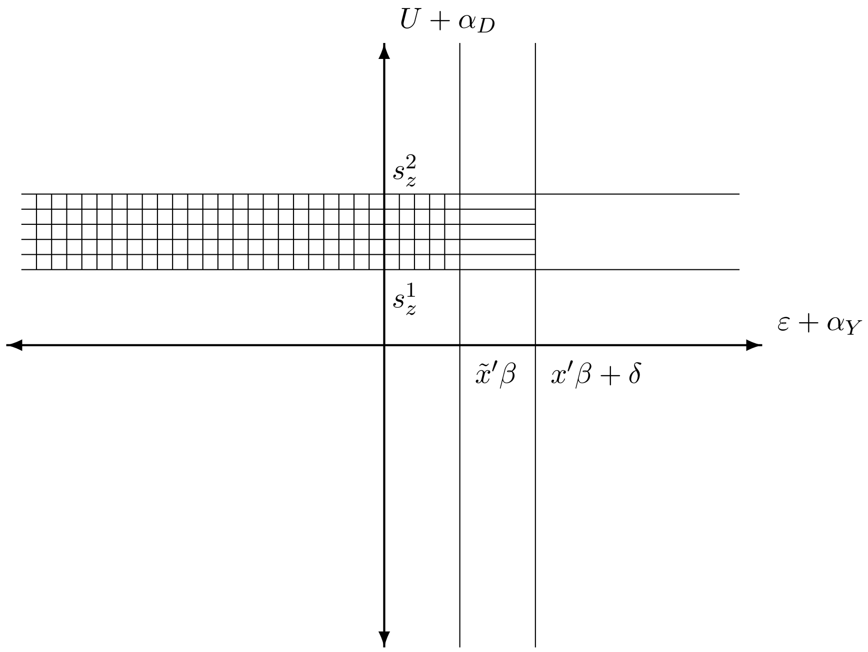
Next, consider two observations, one with characteristics $(x'\beta, z'\gamma)$ and the other with characteristics $(\tilde{x}'\beta, z'\gamma)$. Suppose $Z'\gamma$ is changed from an initial level of s_z^1 to a new level, s_z^2, for both of these observations. In the graph below, the increase in the probability that $D = 1$ and $Y = 1$ conditional on the exogenous covariates for the first observation (the one with characteristics equal x) is represented by the horizontally shaded region. Mathematically, this change equals

$$\begin{aligned} &E[DY|X'\beta = x'\beta, Z'\gamma = s_z^2] - E[DY|X'\beta = x'\beta, Z'\gamma = s_z^1] \\ &\qquad = P\big(U \in (s_z^1 + \alpha_D, s_z^2 + \alpha_D]\big)P\big(\varepsilon \leq x'\beta + \delta + \alpha_Y | U \in (s_z^1 + \alpha_D, s_z^2 + \alpha_D]\big). \end{aligned} \tag{4}$$

[7]Note that if $\tilde{\gamma} = c\gamma$, for some $c \in \mathbb{R} \setminus \{0\}$, the sigma algebra generated by $(X', Z'\gamma)$ will be the same as the sigma algebra generated by $(X', Z'\tilde{\gamma})$ so that $E[Y|X, Z'\gamma] = E[Y|X, Z'\tilde{\gamma}]$.

On the other hand, for the second observation, the change in the probability that $D = 0$ and $Y = 1$ conditional on the exogenous variables is represented by the vertically shaded region. We can express the decrease in this probability as

$$\begin{aligned} &E[(1-D)Y|X'\beta = \tilde{x}'\beta, Z'\gamma = s_z^2] - E[DY|X'\beta = \tilde{x}'\beta, Z'\gamma = s_z^1] \\ &\qquad = -P\big(U \in (s_z^1 + \alpha_D, s_z^2 + \alpha_D]\big)P\big(\varepsilon \leq \tilde{x}'\beta + \alpha_Y | U \in (s_z^1 + \alpha_D, s_z^2 + \alpha_D]\big). \end{aligned} \tag{5}$$



By assumption (A-2) these two changes exactly offset each other if and only if $\tilde{x}'\beta = x'\beta + \delta$. Furthermore, since the left hand sides of equations (4) and (5) can be evaluated from the data a.e. with respect to the joint distribution of $(X'\beta, Z'\gamma)$, we can identify such pairs of observations.

Intuitively, if $Z'\gamma$ changes from s_z^1 to s_z^2, and $X'\beta$ remains constant, this change effects Y through a change in D only. Thus, if for two different covariate levels, $x'\beta$ and $\tilde{x}'\beta$, it is the case that

$$\begin{aligned} &E[DY|Z'\gamma = s_z^2, X'\beta = x'\beta] - E[DY|Z'\gamma = s_z^1, X'\beta = x'\beta] \\ &\qquad + E[(1-D)Y|Z'\gamma = s_z^2, X'\beta = \tilde{x}'\beta] - E[(1-D)Y|Z'\gamma = s_z^1, X'\beta = \tilde{x}'\beta] = 0, \end{aligned}$$

then the change in covariate levels from $x'\beta$ to $\tilde{x}'\beta$ must exactly offset the effect of the change in D. Finding such observations, however, requires that the intersection of the supports of $(Z'\gamma, X'\beta)$ and $(Z'\gamma, X'\beta + \delta)$ is nonempty, and that $X'\beta$ can be varied exogenously while holding $Z'\gamma$ constant. These are guaranteed by assumptions (A-3) and (A-4).

Conditional on knowing γ and β (or having a method to consistently estimate them) identification of δ requires a weaker version of assumption (A-3). In particular, δ could be identified and estimated if

$Z'\gamma$ is discrete.[8] The stronger version of assumption (A-3) was made for the estimation of β.

We conclude this section with a remark on scale normalization on the parameters. Let $a_1 > 0, a_2 > 0$. Then

$$\begin{aligned} Y &= 1\{\alpha_Y + X'\beta + \delta D - \varepsilon \geq 0\} = 1\{\alpha_Y a_2 + X'\beta a_2 + \delta a_2 D - \varepsilon a_2 \geq 0\} = 1\{\alpha_{aY} + X'\beta_a + \delta_a D - \varepsilon_a \geq 0\}, \\ D &= 1\{\alpha_D + Z'\gamma - U \geq 0\} = 1\{\alpha_D a_1 + Z'\gamma a_1 - U a_1 \geq 0\} = 1\{\alpha_{aD} + Z'\gamma_a - U_a \geq 0\}. \end{aligned}$$

Moreover,

$$\begin{aligned} &E[DY|Z'\gamma = s_z^2, X'\beta = x'\beta] - E[DY|Z'\gamma = s_z^1, X'\beta = x'\beta] \\ &\qquad + E[(1-D)Y|Z'\gamma = s_z^2, X'\beta = \tilde{x}'\beta] - E[(1-D)Y|Z'\gamma = s_z^1, X'\beta = \tilde{x}'\beta] = 0 \\ &\qquad \Leftrightarrow E[DY|Z'\gamma_a = s_z^2 a_1, X'\beta_a = x'\beta a_2] - \big(E[DY|Z'\gamma_a = s_z^1 a_1, X'\beta_a = x'\beta a_2] \\ &\qquad + E[(1-D)Y|Z'\gamma_a = s_z^2 a_1, X'\beta_a = \tilde{x}'\beta a_2] - E[(1-D)Y|Z'\gamma_a = s_z^1 a_1, X'\beta_a = \tilde{x}'\beta a_2]\big) = 0, \end{aligned}$$

and if one of these equations holds then $\delta = \tilde{x}'\beta - x'\beta \Leftrightarrow \delta a_2 = \tilde{x}'\beta a_2 - x'\beta a_2$. Thus, identification of the coefficient of D is not affected by the scale normalization, and whatever scale normalization is adopted for β is adopted for the outcome equation, and hence, for δ as well.

4.3 Identification of the remaining parameters:

If we assume F_U is symmetric around 0, identification of α_D follows from Chen (1999). If we assume F_ε is symmetric around 0 as well, then identification of α_Y follows from Chen (1999) once we note that $F_\varepsilon(\alpha_Y + t) = E[DY|Z'\gamma = a, X'\beta = t - \delta] + E[(1-D)Y|Z'\gamma = a, X'\beta = t]$. These arguments also show that F_U and F_ε are identified. Finally, note that U and ε are independent if and only if

$$\begin{aligned} F_{U,\varepsilon}(\alpha_D + a, \alpha_Y + b) &= E[DY|Z'\gamma = a, X'\beta = b - \delta] = F_U(\alpha_D + a)\cdot F_\varepsilon(\alpha_Y + b) = E[D|Z'\gamma = a]\cdot F_\varepsilon(\alpha_Y + b) \\ &= E[D|Z'\gamma = a] \cdot \big(E[DY|Z'\gamma = a, X'\beta = b - \delta] + E[(1-D)Y|Z'\gamma = a, X'\beta = b]\big). \end{aligned}$$

Thus, testing whether U and ε are independent can be done even if α_D and α_Y are unknown.

5 Estimation:

5.1 Estimation of γ and β:

To estimate γ we could choose one of several methods available for estimation of the coefficients in linear index models. Here we propose to use the method developed in Powell, Stock and Stoker (1989) (PSS hereafter). The main reason for this is that this method delivers an easy-to-compute estimator

[8] For discrete $X'\beta$, δ can be identified if there exist $s_z^1, s_z^2, s_x^1, s_x^2$ values such that

$$\begin{aligned} &E[DY|X'\beta = s_x^1, Z'\gamma = s_z^2] - E[DY|X'\beta = s_x^1, Z'\gamma = s_z^1] \\ &\qquad + \big(E[(1-D)Y|X'\beta = s_x^2, Z'\gamma = s_z^2] - E[(1-D)Y|X'\beta = s_x^2, Z'\gamma = s_z^1]\big) = 0. \end{aligned}$$

In that case, δ would be identified as $s_x^2 - s_x^1$ just as in the continuous $X'\beta$ case. However, this seems to be a very special situation.

whose properties can be analyzed in a straightforward fashion. An additional advantage of this method is that it uses a weighting scheme which puts low weight on observations that are drawn from parts of the underlying population which have low likelihood. Define:

$$\hat{\tilde{\gamma}} := \frac{-2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{1}{h_{1n}}\right)^{d_z+1} \frac{\partial K_z}{\partial u} \left(\frac{Z_i - Z_j}{h_{1n}}\right) (D_i - D_j),$$

where $K_z(u)$ is a kernel function satisfying the assumptions of PSS. Theorems 3.3 and 3.4 of PSS tell us that if the data generating process satisfies certain regularity conditions and if the kernel function and the bandwidth are suitably chosen then $\sqrt{n}(\hat{\tilde{\gamma}} - \tilde{\gamma})$ will have an approximately normal distribution, where $\tilde{\gamma} = -2E\left(D\frac{\partial f_z(Z)}{\partial z}\right) = E\left[f_z(Z) f_U(\alpha_D + Z'\gamma)\right]\gamma$. Since $|\gamma_{3,d_3}| = 1$,

$$\hat{\gamma} = \frac{\hat{\tilde{\gamma}}}{|\hat{\tilde{\gamma}}_{3,d_3}|}.$$

Since the method developed in PSS yields straightforward estimators for index coefficients, and since β's are index coefficients themselves, a natural approach for estimating β is to try to extend this method to develop a simple estimator for β. As in the previous section, we have

$$\frac{\partial E[Y|Z'\tilde{\gamma} = t, X = x]}{\partial x} = \left[\frac{\partial P(\tilde{U} \leq \alpha_D + t, \varepsilon \leq \alpha_Y + x'\beta + \delta)}{\partial(x'\beta)} + \frac{\partial P(\tilde{U} > \alpha_D + t, \varepsilon \leq \alpha_Y + x'\beta)}{\partial(x'\beta)}\right]\beta,$$

where $\tilde{U} = E\left[f_z(Z) f_U(\alpha_D + Z'\gamma)\right]U = \tilde{\gamma}_{3,d_3}U$. The vector $\partial E[Y|Z'\tilde{\gamma} = t, X = x]/\partial x$ will remain to be proportional to the vector β if we multiply it by a positive weight which may depend on the point (x, t). The density, $f_{x,z\tilde{\gamma}}$, of $(X, Z'\tilde{\gamma})$ will prove to be a convenient weighting function. These arguments indicate that the methods of PSS are applicable to the estimation of β; however, there is one important distinction here: since we do not know the value of $\tilde{\gamma}$, we need to replace it by an estimated value, and when we construct our estimator of β we have to take this fact into account. Thus, in the following, we follow the same steps as in PSS to derive an estimator. The estimator we get as a result of this process will be infeasible. To make this estimator feasible, we then replace the unknown $\tilde{\gamma}$ parameter with its estimated value from the first stage.

To proceed with the estimation of β, let $\psi(x, t) := E[Y|X = x, Z'\tilde{\gamma} = t]$ and $d_x := d_1 + d_2$. Lemma 2.1 of PSS implies that under assumptions (C-1)-(C-3) given in Appendix (B.1),

$$\tilde{\beta} := E\left(f_{x,z'\tilde{\gamma}} \frac{\partial \psi}{\partial x}\right) = -2E\left(Y \frac{\partial f_{x,z'\tilde{\gamma}}}{\partial x}\right).$$

We can use the sample version of the last expression to estimate β. Picking a symmetric kernel and using the kernel density estimator to estimate $f_{x,z'\tilde{\gamma}}$, we propose the following as our estimator for β:

$$\hat{\tilde{\beta}}^{(F)} := \frac{-2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{1}{h_{2n}}\right)^{d_x+2} \frac{\partial K_{x,t}}{\partial x} \left(\frac{(X_i', Z_i'\hat{\gamma})' - (X_j, Z_j'\hat{\gamma})'}{h_{2n}}\right) [Y_i - Y_j], \tag{6}$$

where

$$\hat{\tilde{\beta}}^{(Inf)} := \frac{-2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\frac{1}{h_{2n}}\right)^{d_x+2} \frac{\partial K_{x,t}}{\partial x} \left(\frac{(X_i', Z_i'\tilde{\gamma})' - (X_j, Z_j'\tilde{\gamma})'}{h_{2n}}\right) [Y_i - Y_j]. \tag{7}$$

To analyze the asymptotic behavior of this estimator, consider

$$\sqrt{n}\left(\hat{\tilde{\beta}}^{(F)} - \tilde{\beta}\right) = \sqrt{n}\left(\hat{\tilde{\beta}}^{(F)} - \hat{\tilde{\beta}}^{(Inf)}\right) + \sqrt{n}\left(\hat{\tilde{\beta}}^{(Inf)} - \tilde{\beta}\right). \tag{8}$$

From the analysis of PSS, we know that the second piece of the expression on the right hand side is asymptotically normal at rate $\sqrt{n}$:

Lemma 5.1 *Suppose assumptions (C-1)-(C-6) given in Appendix (B.1) hold. If h_{2n} obeys $nh_{2n}^{d_x+3} \to \infty$ and $nh_n^{2s_{x,t}} \to 0$ as $n \to \infty$, then*

$$\sqrt{n}(\hat{\tilde{\beta}}^{(Inf)} - \tilde{\beta}) \overset{d}{\Rightarrow} N\left(0, \Sigma_{\tilde{\beta}}\right),$$

where $\Sigma_{\tilde{\beta}} := 4E[r_{\tilde{\beta}}(X_i, Z_i'\tilde{\gamma})r_{\tilde{\beta}}(X_i, Z_i'\tilde{\gamma})'] - 4\tilde{\beta}\tilde{\beta}'$, and

$$r_{\tilde{\beta}}(X_i, Z_i'\tilde{\gamma}) := f_{x,z'\tilde{\gamma}}(X_i, Z_i'\tilde{\gamma})\frac{\partial \psi(X_i, Z_i'\tilde{\gamma})}{\partial x} - [Y_i - \psi(X_i, Z_i'\tilde{\gamma})]\frac{\partial f_{x,z'\tilde{\gamma}}(X_i, Z_i'\tilde{\gamma})}{\partial x}.$$

The above result is a restatement of Theorem 3.3 of Powell, Stock and Stoker (1989). This result, however, gives us only part of the information that we need to analyze the asymptotic behavior of $\hat{\tilde{\beta}}^{(F)}$. To understand the asymptotic behavior of $\hat{\tilde{\beta}}^{(F)}$ fully we also need to study the asymptotic behavior of $\sqrt{n}\left(\hat{\tilde{\beta}}^{(F)} - \hat{\tilde{\beta}}^{(Inf)}\right)$. Using the definitions of $\hat{\tilde{\beta}}^{(F)}$ and $\hat{\tilde{\beta}}^{(Inf)}$, we can write this term as

$$\sqrt{n}\left(\hat{\tilde{\beta}}^{(F)} - \hat{\tilde{\beta}}^{(Inf)}\right) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \frac{-2(Y_i - Y_j)}{\sqrt{n}(n-1)h_{2n}^{d_x+2}}\left\{\frac{\partial K_{x,t}}{\partial x}\left(\frac{(X_i', Z_i'\hat{\tilde{\gamma}})' - (X_j, Z_j'\hat{\tilde{\gamma}})'}{h_{2n}}\right) - \frac{\partial K_{x,t}}{\partial x}\left(\frac{(X_i', Z_i'\tilde{\gamma})' - (X_j, Z_j'\tilde{\gamma})'}{h_{2n}}\right)\right\}.$$

Assuming that $K_{x,t}(x,t) = K_x(x)K_t(t)$,

$$\sqrt{n}\left(\hat{\tilde{\beta}}^{(F)} - \hat{\tilde{\beta}}^{(Inf)}\right) = \textstyle\sum_{i,j\neq i} \frac{-2Y_i}{\sqrt{n}(n-1)h_{2n}^{d_x+2}}\frac{\partial K_x}{\partial x}\left(\frac{X_i - X_j}{h_{2n}}\right)\left\{K_t\left(\frac{Z_i'\hat{\tilde{\gamma}} - Z_j'\hat{\tilde{\gamma}}}{h_{2n}}\right) - K_t\left(\frac{Z_i'\tilde{\gamma} - Z_j'\tilde{\gamma}}{h_{2n}}\right)\right\}.$$

If K_t is continuously differentiable the Mean Value Theorem implies that this last expression equals

$$\sum_{l=1}^{d_z}\sqrt{n}\left(\hat{\tilde{\gamma}}_l - \tilde{\gamma}_l\right)\sum_{i,j\neq i}\frac{-2(Z_{il} - Z_{jl})Y_i}{n(n-1)h_{2n}^{d_x+3}}\frac{\partial K_x}{\partial x}\left(\frac{X_i - X_j}{h_{2n}}\right)K_t'\left(\frac{Z_i'\hat{\xi} - Z_j'\hat{\xi}}{h_{2n}}\right), \tag{9}$$

for some $\hat{\xi}$ between $\hat{\tilde{\gamma}}$ and $\tilde{\gamma}$.

In Appendix (B) we analyze the asymptotic behavior of

$$\sum_i\sum_{j\neq i}\frac{-2(Z_{il} - Z_{jl})Y_i}{n(n-1)h_{2n}^{d_x+3}}\frac{\partial K_x}{\partial x}\left(\frac{X_i - X_j}{h_{2n}}\right)K_t'\left(\frac{Z_i'\hat{\xi} - Z_j'\hat{\xi}}{h_{2n}}\right).$$

That analysis is done under the following assumptions:

(C-7) $K_{xt}(x, z'\tilde{\gamma}) = K_x(x)K_t(z'\tilde{\gamma})$. The first and second derivatives of K_{xt} are bounded.

(C-8) Define, $\tilde{s}_{xt} := \max\{5, s_{xt}\}$. Replace s_{xt} in assumptions [(C-5)] and [(C-6)] by $\tilde{s}_{xt}$. Assume that $nh_{2n}^8 \to \infty$, and $nh_{2n}^{2\tilde{s}_{xt}} \to 0$.

(C-9) Define $\varphi_2(x,t) := E(||Z||^2|X = x, Z'\tilde{\gamma} = t)$. There exists an integrable function $\tilde{m}_2(x,t)$ such that,

$$\left|\left|\varphi_2(x+s_x, t+s_t)f_{xz'\tilde{\gamma}}(x+s_x, t+s_t) - \varphi_2(x,t)f_{xz'\tilde{\gamma}}(x,t)\right|\right| < \tilde{m}_2(x,t)||s||.$$

Assumption (c-8) requires that $nh_{2n}^8 \to \infty$ even when $d_x = 1$, and this is stringent. This could be relaxed to requiring $nh_{2n}^5 \to \infty$ at the cost of assuming $E\|Z\|^3 < \infty$ and imposing a Lipschitz condition on $E(||Z||^3|X = x, Z'\tilde{\gamma} = t)$.

Lemma 5.2 *Under assumptions (C-1) through (C-9), for each* $l = \{1, ..., d_z\}$

$$\sum_i \sum_{j \neq i} \frac{-2(Z_{il} - Z_{jl})Y_i}{n(n-1)h_{2n}^{d_x+3}} \frac{\partial K_x}{\partial x}\left(\frac{X_i - X_j}{h_{2n}}\right) K_t'\left(\frac{Z_i'\hat{\xi} - Z_j'\hat{\xi}}{h_{2n}}\right)$$
$$= \sum_i \sum_{j \neq i} \frac{-2(Z_{il} - Z_{jl})Y_i}{n(n-1)h_{2n}^{d_x+3}} \frac{\partial K_x}{\partial x}\left(\frac{X_i - X_j}{h_{2n}}\right) K_t'\left(\frac{Z_i'\tilde{\gamma} - Z_j'\tilde{\gamma}}{h_{2n}}\right) + o_P(1) =: \hat{C}_l + o_P(1). \quad (10)$$

The proof of this theorem is given in Appendix (B). Next, we turn to the double summation on the right hand side of (10). For the moment suppose that this term has finite expectation. To analyze this term we will appeal to Lemma 3.1 of PSS. To make this application clearer it is helpful to first add and subtract its expectation and define

$$p_n^{\beta\gamma}(i,j) = \frac{(Z_{il} - Z_{jl})(Y_i - Y_j)}{h_{2n}^{d_x+2}} \frac{\partial K_x}{\partial x}\left(\frac{X_i - X_j}{h_{2n}}\right) K_t'\left(\frac{Z_i'\tilde{\gamma} - Z_j'\tilde{\gamma}}{h_{2n}}\right).$$

so that the last summation in (5.2) equals

$$-\frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \{p_n^{\beta\gamma}(i,j) - E[p_n^{\beta\gamma}(i,j)]\} - 2E[p_n^{\beta\gamma}(i,j)]. \quad (11)$$

To verify that $E[||p_n^{\beta}(i,j)||^2] = o(n)$, recall that $\varphi_2(X_i, Z_i'\tilde{\gamma}) := E(||Z_i - Z_j||^2|X_i, Z_i'\tilde{\gamma})$. Then

$$E\left|\left|\frac{(Z_{il} - Z_{jl})(Y_i - Y_j)}{h_{2n}^{d_x+2}} \frac{\partial K_x}{\partial x}\left(\frac{X_i - X_j}{h_{2n}}\right) K_t'\left(\frac{Z_i'\tilde{\gamma} - Z_j'\tilde{\gamma}}{h_{2n}}\right)\right|\right|^2 \leq$$
$$\int\int \left|\left|\frac{\partial K_x}{\partial x}\left(\frac{X_i - X_j}{h_{2n}}\right)\right|\right|^2 \frac{\varphi_2(X_i, Z_i'\tilde{\gamma})}{h_{2n}^{2d_x+4}} \left[K_t'\left(\frac{Z_i'\tilde{\gamma} - Z_j'\tilde{\gamma}}{h_{2n}}\right)\right]^2$$
$$\times f_{xz'\tilde{\gamma}}(X_i, Z_i'\tilde{\gamma})f_{xz'\tilde{\gamma}}(X_j, Z_j'\tilde{\gamma})\mathrm{d}X_i\mathrm{d}(Z_i'\tilde{\gamma})\mathrm{d}X_j\mathrm{d}(Z_j'\tilde{\gamma}) = O(h_{2n}^{d_x+3}). \quad (12)$$

Since $nh_{2n}^{d_x+3} \to \infty$, the condition required for the application of the PSS lemma is satisfied. Thus, the first term in (11) equals

$$-\frac{2}{n}\sum_{i=1}^{n}\left\{E\left[p_n^{\beta\gamma}(i,j)|X_i,Z_i,D_i,Y_i\right]-E\left[p_n^{\beta\gamma}(i,j)\right]\right\}+o_P(1). \tag{13}$$

To deal with this last summation, we appeal to Chebyshev's Law of Large Numbers. To apply this result, we need to verify that the sum of the variances of all the terms is $o(n^2)$. Each term of the summation has mean 0. Inequality (12) implies that the variance of each term in the summation is $O(h_{2n}^{d_x+3})$. Therefore, the condition of the Chebyshev's Law of Large Numbers is satisfied and (13) is $o_P(1)$.

So far we assumed that

$$E\left[\frac{(Z_{il}-Z_{jl})Y_i}{h_{2n}^{d_x+3}}\frac{\partial K_x}{\partial x}\left(\frac{X_i-X_j}{h_{2n}}\right)K_t'\left(\frac{Z_i'\tilde{\gamma}-Z_j'\tilde{\gamma}}{h_{2n}}\right)\right] \tag{14}$$

exists and remains finite as $n\to\infty$. As shown in the Supplementary Appendix, the following assumption guarantees this is indeed the case:

(C-10) Define $\kappa_l(x_1,t_1) := E\big(Z_{1l}|X_1=x_1, Z_1'\tilde{\gamma}=t_1\big)$. We require that for each $(q_1,q_2)\in\mathbb{Z}_+^2$, such that $q_1+q_2=2$, the following Lipschitz conditions are satisfied: For some integrable function $\tilde{m}_{\kappa_l}(x,t)$, $\tilde{m}_{f,2}(x,t)$

$$\left|\left|\frac{\partial^2\kappa_l(x+s_x,t+s_t)}{\partial x^{q_1}\partial(z'\tilde{\gamma})^{q_2}}-\frac{\partial^2\kappa_l(x,t)}{\partial x^{q_1}\partial(z'\tilde{\gamma})^{q_2}}\right|\right|\le \tilde{m}_{\kappa_l}(x,t)||s||,$$
$$\left|\left|\frac{\partial^2 f_{xz'\tilde{\gamma}}(x+s_x,t+s_t)}{\partial x^{q_1}\partial(z'\tilde{\gamma})^{q_2}}-\frac{\partial^2 f_{xz'\tilde{\gamma}}(x,t)}{\partial x^{q_1}\partial(z'\tilde{\gamma})^{q_2}}\right|\right|\le \tilde{m}_{f,2}(x,t)||s||,$$

with $E\big[\big(1+|\kappa_l(X,Z'\tilde{\gamma})|+\|(\frac{\partial\kappa_l(X,Z'\tilde{\gamma})}{\partial x},\frac{\partial\kappa_l(X,Z'\tilde{\gamma})}{\partial(z'\tilde{\gamma})})\|+\tilde{m}_{\kappa_l}(X,Z'\tilde{\gamma})\big)\tilde{m}_{f,2}(X,Z'\tilde{\gamma})\big]<\infty$. In addition, each element of the $1\times d_x$ vector $C_l := E\left[\psi\left\{2\kappa_l\frac{\partial^2 f_{xz'\tilde{\gamma}}}{\partial t\partial x}+\frac{\partial\kappa_l}{\partial t}\frac{\partial f_{xz'\tilde{\gamma}}}{\partial x}+\frac{\partial\kappa_l}{\partial x}\frac{\partial f_{xz'\tilde{\gamma}}}{\partial t}+\frac{\partial^2\kappa_l}{\partial t\partial x}f_{xz'\tilde{\gamma}}\right\}\right]$ is finite.

Finally, because the product of an $O_P(1)$ random variable with a random variable that is $o_P(1)$ is $o_P(1)$, is asymptotically equivalent to

$$\sqrt{n}(\hat{\tilde{\beta}}^{(F)}-\hat{\tilde{\beta}}^{(Inf)})=(-2)\sum_{l=1}^{d_z}\sqrt{n}(\hat{\tilde{\gamma}}_l-\tilde{\gamma}_l)C_l+o_P(1).$$

Theorem 5.1 *Under assumptions (C-1)-(C-10),*

$$\sqrt{n}(\hat{\tilde{\beta}}^{(F)}-\tilde{\beta})\overset{d}{\Rightarrow}N\left(0,\Sigma_{\hat{\tilde{\beta}}^{(F)}}\right)$$

with $\Sigma_{\hat{\tilde{\beta}}^{(F)}}=\big(4E[r_{\tilde{\beta}}(X,Z'\tilde{\gamma})'r_{\tilde{\beta}}(X,Z'\tilde{\gamma})]-4\tilde{\beta}\tilde{\beta}'\big)+4C'\big(4E[r_{\tilde{\gamma}}(Z)'r_{\tilde{\gamma}}(Z)]-4\tilde{\gamma}\tilde{\gamma}'\big)C$ $-4\big(E[2r_{\tilde{\beta}}(X,Z'\tilde{\gamma})'2r_{\tilde{\gamma}}(Z)]-4\tilde{\beta}\tilde{\gamma}'\big)C=\Sigma_{\tilde{\beta}}+4C'\Sigma_{\tilde{\gamma}}C-4\Sigma_{\tilde{\beta}\tilde{\gamma}}C$, *where* $\Sigma_{\tilde{\beta}\tilde{\gamma}}$ *is the asymptotic covariance of* $\sqrt{n}(\hat{\tilde{\beta}}^{(Inf)}-\tilde{\beta})$ *and* $\sqrt{n}(\hat{\tilde{\gamma}}-\tilde{\gamma})$, *and* C *is the* $d_z\times d_x$ *matrix whose* l^{th}*-row equals* C_l, *where* C_l *is as above.*

The proof of this result is given in Appendix (B). The proof uses results developed in this section as well as results from PSS. In the same Appendix, we also discuss how $\Sigma_{\hat{\hat{\beta}}(F)}$ can be consistently estimated. Finally, define

$$\hat{\beta} := \frac{\hat{\hat{\beta}}^F}{|\hat{\hat{\beta}}^F_{k^*}|}.$$

5.2 Estimation of δ:

Having shown that we can identify δ under our assumptions, we move on to its estimation. We will break this task into smaller steps by first considering an infeasible estimator for δ and later developing an estimator which takes into account that γ, β and the conditional expectation functions that are part of the infeasible estimator are not exactly known, but are estimated.

By assumption (A-3), $Z'\gamma$ is a continuous random variable. For the estimation of δ it will be more convenient to work with the derivatives of $E[DY|X'\beta = s_x, Z'\gamma = s_z]$ and $E[(1-D)Y|X'\beta = s_x, Z'\gamma = s_z]$. Under assumptions (A-1)-(A-3) these conditional expectation functions are differentiable in (s_x, s_z), and their derivatives are also identified a.e. with respect to the joint distribution of $(X'\beta, Z'\gamma)$. Define

$$\begin{aligned} g_1(s_z, s_x) &:= \frac{\partial}{\partial s_z} E[DY|Z'\gamma = s_z, X'\beta = s_x] = \int_{-\infty}^{s_x+\delta} f_{U,\varepsilon}(s_z, e)\mathrm{d}e, \\ g_0(s_z, s_x) &:= \frac{\partial}{\partial s_z} E[(1-D)Y|Z'\gamma = s_z, X'\beta = s_x] = -\int_{-\infty}^{s_x} f_{U,\varepsilon}(s_z, e)\mathrm{d}e. \end{aligned}$$

Our analysis above tells us that for any two observations i, j such that $Z_i'\gamma = Z_j'\gamma$, and

$$\frac{\partial}{\partial s_z} E[D_i Y_i | X_i'\beta, Z_i'\gamma] + \frac{\partial}{\partial s_z} E[(1-D_j)Y_j | X_j'\beta, Z_j'\gamma] = 0,$$

we have $X_j'\beta - X_i'\beta = \delta$. As long as $P\big((Z'\gamma, X'\beta) \in \mathcal{S}\big) > 0$, with $\mathcal{S} := \mathrm{Supp}(Z'\gamma, X'\beta) \cap \mathrm{Supp}(Z'\gamma, X'\beta + \delta)$, we will find such pairs of observations in our data with positive probability.

This identification information cannot be directly implemented as an estimation procedure. In the first place, the parameters γ, β, g_1 and g_0 are unknown, so even if pairs of observations for which $Z_i'\gamma = Z_j'\gamma$ and $g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta) = 0$ existed in the sample, we would not be able to determine precisely which pairs these were. Nevertheless, since

$$\tilde{g}(u, a) := \int_{-\infty}^{a} f_{U,\varepsilon}(u, s)\mathrm{d}s$$

is continuous in u, and $\tilde{g}^{-1}(u, g)$ is continuous for each u in its second argument, where $\tilde{g}^{-1}$ denotes the inverse of the $\tilde{g}(u, g)$ function with respect to its second argument[9], $X_j'\beta - X_i'\beta$ will approximately be δ whenever $Z_i'\gamma$ approximately equals $Z_j'\gamma$, and

$$\frac{\partial}{\partial s_z} E[D_i Y_i | X_i'\beta, Z_i'\gamma] + \frac{\partial}{\partial s_z} E[(1-D_j)Y_j | X_j'\beta, Z_j'\gamma]$$

[9] In other words, $\tilde{g}^{-1}(u, \tilde{g}(u, a)) = a$.

is approximately 0. To operationalize this idea, we need to specify precisely which approximations are satisfactory. On the other hand, once we consider the differences in linear indices, that is $(X_j'\beta - X_i'\beta)$'s, for pairs of observations where $Z_j'\gamma$ approximately equals $Z_i'\gamma$ and $g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)$ is approximately 0 as acceptable, the possibility of having multiple such pairs of observations arises. We resolve this issue by taking a weighted average of $X_j'\beta - X_i'\beta$, where the average is taken over pairs of observations for which $Z_i'\gamma \approx Z_j'\gamma$ and $g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta) \approx 0$,[10] and the weights decline as the sample size or the distance between $\left(Z_i'\gamma, g_1(Z_i'\gamma, X_i'\beta)\right)$ and $\left(Z_j'\gamma, -g_0(Z_j'\gamma, X_j'\beta)\right)$ gets large. A convenient algebraic form for such a weight function is a kernel weight. The analysis thus far suggests the following infeasible estimator:

$$\hat{\delta}^{(Inf)} := \frac{\sum_i \sum_j \frac{1}{h_{3n}^2} k\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right)(X_j - X_i)'\beta}{\sum_i \sum_j \frac{1}{h_{3n}^2} k\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma X_j'\beta)}{h_{3n}}\right)}, \tag{15}$$

where k is a known smooth function that integrates to 1, and $\{h_{3n}\}$ is a sequence of bandwidths which tends to 0 as the sample size increases.

The numerator of this expression is a weighted sum of the differences of linear indices of observations for which $\left(Z_i'\gamma, g_1(Z_i'\gamma, X_i'\beta)\right)$ and $\left(Z_j'\gamma, -g_0(Z_j'\gamma, X_j'\beta)\right)$ are close. Since the kernel function $k(\cdot)$ necessarily declines to 0 as either of its arguments tends to infinity in magnitude, any pair of observations for which $\left(Z_i'\gamma, g_1(Z_i'\gamma, X_i'\beta)\right) \neq \left(Z_j'\gamma, -g_0(Z_j'\gamma, X_j'\beta)\right)$ will necessarily receive declining weight as the sample size grows and the bandwidth shrinks to 0. At the same time, as we noted earlier, there might be multiple pairs of observations for which $k\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right) \neq 0$, and when this happens the numerator does not estimate δ, but a multiple of it. The denominator of the proposed estimator is there to solve this problem.

As we will show later, the probability limit of the denominator equals

$$E[f_{z'\gamma, x'\beta}(Z'\gamma, X'\beta + \delta)].$$

If $\mathcal{S}$ has measure 0, this expectation will be zero. Identification of δ is made possible by observations coming from the set where the supports of $(Z'\gamma, X'\beta)$ and $(Z'\gamma, X'\beta + \delta)$ overlap. If this set has low probability, the likelihood of having observations from this set will be small, and consequently any reasonable estimator of δ that is based on the identification result presented here will tend to have a large variance. This fact is reflected in our estimator through the inverse relation between the asymptotic variance of our estimator and the probability of this set. In particular, if this set has zero probability, the variance of our estimator will be infinite.

An alternative way of looking at the proposed estimator is to note that it equals the kernel regression estimator of the conditional expectation of $X_j'\beta - X_i'\beta$ given $Z_j'\gamma - Z_i'\gamma = 0$ and $g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta) = 0$. This interpretation of the estimator may help us better understand its asymptotic behavior.

As we already noted, the estimator proposed above is infeasible, because in reality we do not know the values of γ, β, g_1 and g_0. Thus, to have a feasible estimator, we need to replace these with their estimated counterparts. In the previous subsection, we devised estimators for γ and β. In addition, in the appendix, we state assumptions under which local polynomial regression estimators of g_1 and g_0

[10] We use $\approx$ to mean "approximately equal".

are consistent. Note that to make these estimated functions uniformly consistent we need to trim out those observations of $(Z'\gamma, X'\beta)$ for which the value of the density $f_{z'\gamma,x'\beta}$ is low. Thus, the feasible estimator we propose is

$$\hat{\delta}^{(F)} := \frac{\sum_i \sum_j \frac{1}{h_{3n}^2} k\left(\frac{(Z_i - Z_j)'\hat{\gamma}, \hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})}{h_{3n}}\right)(X_j - X_i)'\hat{\beta}\hat{I}_i\hat{I}_j}{\sum_i \sum_j \frac{1}{h_{3n}^2} k\left(\frac{(Z_i - Z_j)'\hat{\gamma}, \hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma} X_j'\hat{\beta})}{h_{3n}}\right)\hat{I}_i\hat{I}_j}, \tag{16}$$

where

$$\hat{I} := 1\{\hat{f}_{z'\hat{\gamma},x'\hat{\beta}}(z'\hat{\gamma}, x'\hat{\beta}) \geq q_0, \ \& \ (w_1, w_2, w_3) \in [-T_n, T_n]^d\},$$

q_0 is a pre-specified positive number, $\{T_n\}_{n=1}^{\infty}$ is a sequence of real numbers which goes to infinity at a slow rate, and (w_1, w_2, w_3) denotes the values of distinct components of (x, z) so that $(X'\beta, Z'\gamma) = (W_1'\beta_1 + W_2'\beta_2, W_2'\gamma_2 + W_3'\gamma_3)$.[11] In addition,

$$\hat{f}_{z'\hat{\gamma},x'\hat{\beta}}(z'\hat{\gamma}, x'\hat{\beta}) = \frac{1}{n\tilde{h}_n^2} \sum_{l=1}^{n} \tilde{K}\left(\frac{Z_l'\hat{\gamma} - z'\hat{\gamma}, X_l'\hat{\beta} - x'\hat{\beta}}{\tilde{h}_n}\right),$$

an $\hat{g}_1$ and $\hat{g}_0$ denote local polynomial regression estimators of g_1 and g_0 using the $Z'\hat{\gamma}$ and $X'\hat{\beta}$ as regressors.

When the density of $(Z'\gamma, X'\beta)$ is uniformly continuous, the set, $\{(s_z, s_x) : f_{z'\gamma,x'\beta}(Z'\gamma, X'\beta) \geq q_0\}$ is compact for each $q_0 \geq 0$. Compactness of this set eases the analysis of terms that involve estimation errors associated with g_1 and g_0 functions. To capture the whole set $\mathcal{S}$ we would ideally let $q_0 \downarrow 0$, or use smooth trimming functions, which is left for future research. In analyzing the asymptotic behavior of $\hat{\delta}^{(F)}$ we rely on Mean Value Theorem and the fact that $\hat{\gamma}$ and $\hat{\beta}$ are consistent for their population counterparts and are $\sqrt{n}$-normal. When we use the Mean Value Theorem we end up with components of W_i, W_j in our summations. Restricting W to lie in d dimensional rectangle of size T_n initially and then letting T_n go to ∞ greatly simplifies the asymptotic analysis. If T_1 is chosen to be a very large number then this approach will have a negligible effect on our estimator even in small samples. In addition, this way we can avoid assuming that W has bounded support.

To derive the asymptotic distribution of $\sqrt{n}\left(\hat{\delta}^{(F)} - \delta\right)$ we impose some regularity conditions in addition to the identification assumptions we have imposed so far. For identification we had to assume probability of the set $\mathcal{S}$ is larger than 0. For estimation we had to use trimming functions. Thus the identification assumption has to be strengthened to

Assumption 5.1 *$P\big((Z'\gamma, X'\beta) \in \mathcal{S} \cap \mathcal{T}\big) > 0$, where $\mathcal{S}$ is as previously defined, and $\mathcal{T} = \{(s_z, s_x) : f_{z'\gamma,x'\beta}(s_z, s_x) \geq q_0, \ f_{z'\gamma,x'\beta}(s_z, s_x + \delta) \geq q_0\}$.*

Analysis of the asymptotic behavior for the estimator for δ will be done assuming we have well behaved estimators of γ and β in the following sense:

[11] Note that if (W_1, W_2, W_3) have a joint density and the matrix $\begin{bmatrix} \beta_1' & \beta_2' & 0 \\ 0 & \gamma_1' & \gamma_2' \end{bmatrix}$ has rank 2, then $(X'\beta, Z'\gamma)$ will also have a joint density. Moreover, if $(\beta_1', \beta_2', 0)'$ and $(0, \gamma_1', \gamma_2')'$ are linearly independent then so will $(\hat{\beta}_1', \hat{\beta}_2', 0)$ and $(0, \hat{\gamma}_1', \hat{\gamma}_2')$ for sufficiently large n with probability close to 1.

Assumption 5.2 *We have estimators, $\hat{\gamma}$ and $\hat{\beta}$, of γ and β respectively, such that $\sqrt{n}(\hat{\gamma} - \gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_i^{\gamma} + o_P(1)$ and $\sqrt{n}(\hat{\beta} - \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_i^{\gamma} + o_P(1)$.*

The analysis in the previous sections shows that this assumption holds under some mild regularity conditions. The next assumption imposes regularity conditions on the density of $(Z'\gamma, X'\beta)$, which we use to ensure that using trimming functions based on $\hat{f}_{z'\hat{\gamma},x'\hat{\beta}}(z'\hat{\gamma}, x'\hat{\beta})$ as opposed to $f_{z'\gamma,x'\beta}(z'\gamma, x'\beta)$ has no effect on the asymptotic distribution of $\hat{\delta}^{(F)}$.

Assumption 5.3

(a) $f_{z'\gamma,x'\beta}$ *is uniformly continuous and bounded on* $\mathbb{R}^2$.

(b) $\tilde{K}$ *is a compactly supported differentiable function (w.l.o.g. its support can be assumed to be contained in the unit cube of $\mathbb{R}^2$) with $\int_{\mathbb{R}^2} K(u)\mathrm{d}u = 1$. Moreover, $\tilde{K}$ is in the linear span of functions $v \geq 0$ satisfying the following property: the subgraph of v, $\{(u,t) : v(u) \geq t\}$, can be represented as a finite number of Boolean operations among sets of the form $\{(u,t) : p(s,t) \geq \varphi(t)\}$, where p is a polynomial on $\mathbb{R}^2 \times \mathbb{R}$ and φ is an arbitrary real function.*

(c) *Partial derivatives of $f_{z'\gamma,x'\beta}$ up to order 3 exist, the third order partial derivatives are Hölder continuous, and $\int_{\mathbb{R}^2} u_z^p u_x^q K(u)\mathrm{d}u = 0$ for $p, q \geq 0$ and $1 \leq p + q \leq 2$.*

(d) *(i) $\tilde{h}_n \downarrow 0$, (ii) $\frac{n\tilde{h}_n^2}{|\log \tilde{h}_n|} \to \infty$, (iii) $\frac{|\log \tilde{h}_n|}{\log\log n} \to \infty$, (iv) $T_n \uparrow \infty$, (v) $\frac{T_n}{\sqrt{n}\tilde{h}_n^3} \to 0$ and (vi) $\sqrt{n}\tilde{h}_n^4 \downarrow 0$.*

(e) *$q_0 > 0$, and for each $(s_z, s_x) \in \mathcal{R} := \{(s_z', s_x') : f_{z'\gamma,x'\beta}(s_z', s_x') = q_0\} \neq \emptyset$, $\left\| \left(\frac{\partial f_{z'\gamma,x'\beta}(s_z,s_x)}{\partial s_z}, \frac{\partial f_{z'\gamma,x'\beta}(s_z,s_x)}{\partial s_x} \right) \right\| > 0$.*

Even though the condition that $\tilde{K}$ is in the linear span of certain functions "seems awkward it is quite general." For example it "is satisfied by $\tilde{K}(u) = \phi(p(u))$, p being a polynomial and ϕ" a continuous, compactly supported, real function, or "if the graph of $\tilde{K}$ is a pyramid (truncated or not), or if $\tilde{K} = 1\{[-1,1]^2\}$." (p. 911 of Gine and Guillou (2002).) Under assumptions 5.3a, b, d Theorem 3.3 of Gine and Guillou (2002) implies that

$$\sqrt{\frac{n\tilde{h}_n^2}{2\log \tilde{h}_n^{-2}}} \sup_{(z'\gamma,x'\beta)\in\mathbb{R}^2} |\hat{f}_{z'\gamma,x'\beta}(z'\gamma, x'\beta) - f_{z'\gamma,x'\beta}(z'\gamma, x'\beta)| = o_P(1). \tag{17}$$

Also under the same assumptions, by Mean Value Theorem we have

$$\sup_{(z'\gamma,x'\beta)\in\overline{A}^*} \left| \frac{1}{n\tilde{h}_n^2} \sum_{i=1}^{n} \tilde{K}\left(\frac{Z_i'\hat{\gamma} - z'\hat{\gamma}, X_i'\hat{\beta} - x'\hat{\beta}}{\tilde{h}_n} \right) - \hat{f}_{z'\gamma,x'\beta}(z'\gamma, x'\beta) \right| = o_P(1), \tag{18}$$

where $\overline{A}^* := \{(s_z, s_x) \in \mathbb{R}^d : f_{Z'\gamma,X'\beta}(s_z, s_x) \geq q_0 - \epsilon_f^*\}$, with $\epsilon_f^* > 0$ chosen such that for each (s_z, s_x) with $f_{z'\gamma,x'\beta}(s_z', s_x') \in [q_0 - \epsilon_f^*, q_0 + \epsilon_f^*]$, $\left\| \left(\frac{\partial f_{z'\gamma,x'\beta}(s_z,s_x)}{\partial s_z}, \frac{\partial f_{z'\gamma,x'\beta}(s_z,s_x)}{\partial s_x} \right) \right\| > 0$.

Next, we impose regularity conditions that help us control the asymptotic behavior of $\hat{g}_1(Z'\hat{\gamma}, X'\hat{\beta})$ and $\hat{g}_0(Z'\hat{\gamma}, X'\hat{\beta})$. Here we use local polynomial regression estimators for these functions, but obviously other estimators for these functions could be used as well. Note that for $r = 0, 1$

$$\hat{g}_r(Z'\hat{\gamma}, X'\hat{\beta}) - \hat{g}_r(Z'\gamma, X'\beta) + \hat{g}_r(Z'\gamma, X'\beta) - g_r(Z'\gamma, X'\beta).$$

To control the behavior of the first term we are going to rely on Mean Value Theorem. To control the asymptotic behavior of the second term above we need to impose conditions so that the local polynomial estimators are well behaved.

Assumption 5.4 *For $r = 0, 1$,*

(a) $\{h_{g_r n}\}$ *satisfies* $nh_{g_r n}^3 / \log n \to \infty$ *and* $nh_{g_r n}^{2(\overline{p}_{g_r}-1)} \to c_{g_r} < \infty$ *for some* $c_{g_r} \geq 0$*, and* $\overline{p}_{g_r} > 3$.

(b) Kernel function $K^{g_r}(\cdot)$ *is a symmetric, continuously differentiable and compactly supported function with Hölder continuous derivative. It has moments of order* $p+1$ *through* $\overline{p}_{g_r} - 1$ *that are equal to zero.*

(c) $P(D = 1) \in (0, 1)$.

(d) The probability density function $f_{U,\varepsilon}$ *is four times continuously differentiable.* [12]

Under these assumptions, following arguments similar to those used in the proof of Theorem 3 of Heckman, Ichimura and Todd (1998), we can show that for $r = 0, 1$, $\hat{g}_r(s_z, s_x)$ is asymptotically linear with trimming:

$$\begin{aligned}\big[\hat{g}_r(z'\gamma, x'\beta) - g_r(z'\gamma, x'\beta)\big] I(z'\gamma, x'\beta) &= n^{-1}\sum_{j=1}^{n} \psi_{ng_r}(D_j, Y_j, Z_j'\gamma, X_j'\beta; z'\gamma, x'\beta) \\ &+ \hat{b}_{\hat{g}_r}(z'\gamma, x'\beta) + \hat{R}_{\hat{g}_r}(z'\gamma, x'\beta),\end{aligned}$$

with $\text{plim}_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{b}_{\hat{g}_r}(Z_j'\gamma, X_j'\beta) = b_{gr} < \infty$, $\text{plim}_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{R}_{\hat{g}_r}(Z_j'\gamma, X_j'\beta) = 0$, and $E[\psi_{ng_r}(D_j, Y_j, Z_j'\gamma, X_j'\beta; z'\gamma, x'\beta)|X'\beta, Z'\gamma] = 0$. We analyze the asymptotic behavior of numerator of $\hat{\delta}^{(F)}$ in multiple steps by replacing estimated quantities with their population counterparts. For example, in one of the steps, we study the asymptotic behavior of

$$\textstyle\sum_i \sum_j \frac{(X_j - X_i)'\hat{\beta}\hat{I}_i\hat{I}_j}{\sqrt{n}(n-1)h_{3n}^2} \Big[k\Big(\frac{(Z_i-Z_j)'\hat{\gamma}, \hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})}{h_{3n}}\Big) - k\Big(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + \hat{g}_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\Big)\Big].$$

We assume that $\hat{\gamma}$ and $\hat{\beta}$ are asymptotically linear at rate $\sqrt{n}$. In addition, under Assumption (5.3) we know the uniform convergence rate of the estimated density to the actual one. In addition, part (e) of that assumption allows us to convert this convergence rate to the convergence rate of the estimated trimming function to its population counterpart. To deal with the term above we also need to know something about how fast $\hat{g}_r(z'\gamma, x'\beta)$ converges to $g_r(z'\gamma, x'\beta)$. For this purpose we assume:

Assumption 5.5 $\{a_n\}_{n=1}^{\infty}$ *is an increasing sequence of numbers diverging to infinity and satisfies* $\frac{\sqrt{n}}{a_n^2 h_{3n}^4} \to c < \infty$, $\frac{nh_{g_r n}^4}{a_n^2 \log n} \to \infty$, *and* $\frac{T_n a_n}{\sqrt{n}} \to 0$.

[12] What we really need is that $\tilde{g}^{-1}$ is $\tilde{p}_g$ times continuously differentiable with $\tilde{p}_g \geq 5$. But using the Implicit Function Theorem, we can write $\frac{\partial \tilde{g}^{-1}(s_z, g)}{\partial g} = \frac{1}{f_{U\varepsilon}(s_z, \tilde{g}^{-1}(s_z, g))}$ and $\frac{\partial \tilde{g}^{-1}(s_z, g)}{\partial s_z} = \frac{\int_{-\infty}^{\tilde{g}^{-1}(s_z, g)} \frac{\partial f_{U\varepsilon}(s_z, t)}{\partial s_z} \mathrm{d}t}{f_{U\varepsilon}(s_z, \tilde{g}^{-1}(s_z, g))}$. Thus our assumption suffices for the desired condition.

Using the asymptotic linearity of $\hat{g}_r$ and modifying the proof of Lemma 5 of Heckman, Ichimura and Todd (1998), we have[13]

$$\sup_{(s_z,s_x)\in\overline{A}^*} a_n|\hat{g}_r(s_z,s_x) - g_r(s_z,s_x)| \xrightarrow{P} 0. \tag{19}$$

Finally, we need to impose some conditions on the kernel k and on the relative convergence rates of $\tilde{h}_n$, h_{3n} and T_n.

Assumption 5.6 *$\sqrt{n}h_{3n}^5 \to 0$, $T_n h_{3n} \to 0$ and $\frac{T_n^2}{nh_{3n}^2\tilde{h}_n^2}\log\left(\frac{1}{\tilde{h}_n^2}\right) \to 0$.*

Assumption 5.7 *k is symmetric, has compact support and is five times continuously differentiable. In addition, $\int k(u_z,u_g)\mathrm{d}u_z\mathrm{d}u_g = 1$, and for $p \geq 1$, $r \geq 1$, $1 \leq p+r \leq 4$, $\int u_z^p u_g^r k(u_z,u_g)\mathrm{d}u_z\mathrm{d}u_g = 0$.*

Theorem 5.2 *Under Assumptions (A-0)-(A-4) and (5.1)-(5.7)*

$$\sqrt{n}\left(\hat{\delta}^{(F)} - \delta\right) = \frac{1}{E[f_{z'\gamma,x'\beta}(Z'\gamma, X'\beta+\delta)1\{(Z'\gamma,X'\beta)\in\mathcal{S}\cap\mathcal{T}\}]}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi^\delta(Z_i'\gamma, X_i'\beta) + b\right] + o_P(1). \tag{20}$$

The proof of this result consists of multiple steps where at each step we analyze the effect of having an estimated quantity, as opposed to its population counterpart. Appendix (A) gives the proof. This proof heavily relies on results presented in Heckman, Ichimura and Todd (1998).

The bias, b, has two terms resulting from the fact that g_1 and g_0 are estimated by local polynomial regression. Specifically, b can be estimated as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n [\hat{\hat{b}}_{g_1}(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{\hat{b}}_{g_0}(Z_i'\hat{\gamma}, X_i'\hat{\beta})]\hat{\lambda}(Z_i'\hat{\gamma}, X_i'\hat{\beta}),$$

where

$$\hat{\lambda}_n(s_z,s_x) := \sum_{l=1}^{n-1}\frac{[X_l'\hat{\beta} - s_x - \delta]I_l}{(n-1)h_{3n}^3}k_2\left(\frac{s_z - Z_l'\gamma, g_1(s_z,s_x) + g_0(Z_l'\hat{\gamma}, X_l'\hat{\beta})}{h_{3n}}\right),$$

and

$$\hat{\hat{b}}_{g_r}(Z_i'\hat{\gamma},X_i'\hat{\beta}) = h_{gn}^{\overline{p}-1}e_2[\hat{M}_{pn}(Z_i'\hat{\gamma},X_i'\hat{\beta})]^{-1}\sum_{k=p+1}^{\overline{p}}\left[\int u^{Q(0)}\cdot u^{Q(k)}\hat{m}_r^{(k)}(Z_i'\hat{\gamma},X_i'\hat{\beta})'\cdot u^{Q(\overline{p}-k)}K_g(u)\mathrm{d}u\right.$$
$$\left.,\ldots,\int u^{Q(p)}\cdot u^{Q(k)}\hat{m}_r^{(k)}(Z_i'\hat{\gamma},X_i'\hat{\beta})'\cdot u^{Q(\overline{p}-k)}K_g(u)\mathrm{d}u\right]\hat{f}_{z'\hat{\gamma},x'\hat{\beta}}^{(\overline{p}-k)}(Z_i'\hat{\gamma},X_i'\hat{\beta})\hat{I}(Z_i'\hat{\gamma},X_i'\hat{\beta}),$$

with $\hat{m}_1^{(k)}(s_z,s_x)$ and $\hat{m}_0^{(k)}(s_z,s_x)$ denoting estimators for k^{th} order partial derivatives of $E[DY|Z'\gamma = s_z, X'\beta = s_x]$ and $E[(1-D)Y|Z'\gamma = s_z, X'\beta = s_x]$ evaluated at (s_z,s_x), respectively, and $\hat{f}_{z'\hat{\gamma},x'\hat{\beta}}^{(\overline{p}-k)}(s_z,s_x)$ denoting an estimator for the vector of $(\overline{p}-k)^{th}$ order partial derivatives of the density of $(Z'\gamma, X'\beta)$ evaluated at (s_z,s_x), and $\hat{M}_{pn}$ and $u^{Q(s)}$ are as on page 284 of Heckman, Ichimura and Todd (1998).

[13] This is shown in the supplementary appendix.

The definition of ψ^δ is given in Appendix A.3. The same Appendix also outlines how the variance of ψ^δ can be estimated. The fact that γ, β, g_0 and g_1 have to be estimated manifests itself in ψ^δ and b. Specifically, $\hat{\gamma}$ appears in three places in the numerator of our estimator, and there are three terms multiplying $\frac{1}{\sqrt{n}}\sum_i \psi_i^\gamma$ reflecting these three sources of additional variability caused by having to estimate γ. Similarly, $\hat{\beta}$ appears in three places in the numerator of our estimator, and, as a result, there are three terms multiplying $\frac{1}{\sqrt{n}}\sum_i \psi_i^\beta$. Then there are two additional terms in ψ^δ representing the additional variability caused by the fact that the g_1 and g_0 functions have to be estimated. The infeasible estimator does not show up in ψ^δ at all. This seems surprising at first, but a closer look at how δ is identified helps us understand this fact. Recall that $\delta = X_j'\beta - X_i'\beta$ if and only if $Z_i'\gamma = Z_i'\gamma$ and $g_{1i} + g_{0j} = 0$. As a reflection of this fact, our infeasible estimator can be thought of as a kernel regression estimator of v, an $n(n-1)\times 1$ dimensional vector, on ω, an $n(n-1)\times 2$ dimensional vector, at $\omega = 0$, with each v_l corresponding to some $X_j'\beta - X_i'\beta$ and each ω_l corresponding to $(Z_i'\gamma - Z_j'\gamma, g_{1i}+g_{0j})$ (with the same i, j). Normally, when one takes the conditional expectation of a random variable given some other random vector there is a residual which is not degenerate, i.e. normally one would have $v = m(\omega) + r$, where $m(\omega) = E(v|\omega)$ and $E(r|\omega) = 0$. In our case, however, $E(v|\omega = 0) = \delta$, which is constant. Thus, there is no remaining variation once conditional expectation of v is taken conditional on $\omega = 0$. So when we kill the bias of the infeasible estimator we also kill its variance.

This paper used estimation procedures that rely on kernels which employ different bandwidths that go to 0 as the sample size increases. The dependence of the estimators on these bandwidths raises the issue of how to select these bandwidths. The existing results on optimal bandwidth selection are not readily applicable to our problem. One reason is because the existing results are not for multiple step estimation methods with this many steps. Second, when one tries to optimize some type of mean squared error with respect to the bandwidth choices then the optimal bandwidth choice often depends on quantities that are unknown. A common way of estimating such quantities is cross validation, which involves replacing these unknown quantities with their corresponding leave-one-out estimators. Since leave-one-out estimator is a random variable, the optimal bandwidth resulting from this procedure will also be a random variable. The results given in our paper, however, are for fixed bandwidth sequences and, as a result, do not exactly cover cross-validation procedure. The third problem we face is that the existing results use some form of integrated squared error as their optimality criterion as these results are often for an estimator of a conditional mean function or its derivatives. Consequently, one might try to choose the bandwidth so that the estimator is close to its population value over the whole range of values of the conditioning variables take in the sample. In our case, the estimator for δ, the main parameter of interest, equals $E[X_j'\beta - X_i'\beta|(X_i'\beta, Z_i'\gamma) \in \mathcal{S}]$, where $\mathcal{S} = \mathrm{Supp}(Z'\gamma, X'\beta) \cap \mathrm{Supp}(Z'\gamma, X'\beta + \delta) = \mathrm{Supp}(Z'\gamma, g_0) \cap \mathrm{Supp}(Z'\gamma, g_1)$. Thus, the criterion function for choosing the optimal bandwidth vector in our case should also depend on the required support condition and will be more complicated than an unconditional integrated square error.

Chen and Vytlacil (2005) studies a nonlinear panel data model with lagged dependent variables. Their two period model with (their) $\delta_2 = 0$ is identical to the model studied here. The aim of their paper is to provide sieve minimum distance and sieve maximum likelihood estimators that are $\sqrt{n}$-consistent, asymptotically normal and efficient under regularity conditions. In its current form, that paper assumes that there is a component in X with $\mathbb{R}$ as its support, so that the identification conditions are converted into simple moment conditions. The paper then uses copula methods to derive an estimator which exploits these moment conditions. We conjecture that the asymptotic variance of our estimator is larger than that of the efficient estimators as we only incorporated into our estimator two out of the three moment conditions that Chen and Vytlacil will develop; our estimator does not use the information contained in the observations for which $Y = 0$. One could modify our estimation method by first considering an alternative estimator that is based on the changes in $E[D(1-Y)|Z'\gamma, X'\beta]$ and $E[(1-D)(1-Y)|Z'\gamma, X'\beta]$ that occur as a result of a change in $Z'\gamma$, and then take an optimally weighted average of our estimator and this alternative estimator. This is left for future research.

6 Conclusion:

This paper proposed a new semiparametric estimation method for estimating binary response models with dummy endogenous variables. In outlining this method, the paper made two contributions. First, it presented an identification result for the problem of estimating binary choice models with an endogenous dummy regressor and also proposed a novel "matching" estimator for the coefficient on the binary endogenous variable in the outcome equation. Second, in order to establish the asymptotic properties of the proposed estimators of the coefficients on the exogenous regressors in the outcome equation the chapter extended the results of PSS to cover the case where the weighted average derivative estimation requires a first step semiparametric procedure.

In studying the estimation of binary response models with binary endogenous variables, here we focused on the case where the binary endogenous regressor has a constant effect in addition to the effect of the exogenous characteristics. A natural extension to this paper would be allowing the binary endogenous regressor to interact with the exogenous regressors, while keeping the linear index structure.

References

[1] Altonji, J. and R. L. Matzkin 2003, "Cross Section and Panel Data Estimators for NonSeparable Models with Endogenous Regressors," manuscript, Yale University.

[2] Altonji, J. and R. L. Matzkin 2005, "Cross Section and Panel Data Estimators for NonSeparable Models with Endogenous Regressors," *Econometrica*, **73**, 1053-1153.

[3] Amemiya, T., 1978, "The Estimation of a Simultaneous Equation Generalized Probit Model," *Econometrica*, **46**, 1193-1205.

[4] Bhattacharya, J., D. McCaffrey, and D. Goldman, 1999, "Estimating Probit Models with Endogenous Covariates," unpublished working paper, RAND.

[5] Blundell, R. and J. Powell, 2004, "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, **71**, 581-593.

[6] Blundell, R. and J. Powell, 2003, "Endogeneity in Nonparametric and Semiparametric Regression Models," in Dewatripont, M., L. P. Hansen, and S. J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Vol II, (Cambridge: Cambridge University Press).

[7] Chen, S., 1999, "Semiparametric Estimation of a Location Parameter in the Binary Choice Model," *Econometric Theory*, **15**, 79-98.

[8] Chen, X. and E. Vytlacil, 2005, "Nonlinear Panel Data Models with Lagged Dependent Variables," manuscript, Stanford University.

[9] Chernozhukov, V. and C. Hansen, 2001, "An IV Model of Quantile Treatment Effects," *Econometrica*, **73**, 245-263.

[10] Chesher, Andrew (2003), "Identification in Nonseparable Models", *Econometrica* 71:5, 1405-1441.

[11] Chesher, Andrew (2005), "Nonparametric Identification under Discrete Variation", *Econometrica*, 73:5 , 1525-1550.

[12] Giné, E. and A. Guillou, (2002), "Rates of Strong Uniform Consistency for Multivariate Kernel Density Estimators", *Annals of H. Poincaré*, **PR-38**, 907-921.

[13] Heckman, J., 1978, "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica* , **46**, 931-959.

[14] Heckman, J., H. Ichimura, and P. Todd, 1998, "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies,*, **65**, 261-294.

[15] Horowitz, J.L. and W. Härdle, 1996, "Direct Semiparametric Estimation of Single Index Models with Discrete Covariates," *Journal of the American Statistical Association*, **91**, 1632-1640.

[16] Ichimura, H., 2004, "Computation of Asymptotic Distribution for Semiparametric GMM Estimators," unpublished manuscript, University College London, CEMMAP and IFS.

[17] Imbens, G. and W. Newey, 2009, "Identification and Estimation of Triangular Simultaneous Equation Models without Additivity," *Econometrica*, **77**, 1481-1512.

[18] Jun, S. J., 2009, "Local Structural Quantile Effects in a Model with a Nonseparable Control Variable," *Journal of Econometrics*, **151**, 82-97.

[19] Lewbel, A., 2000, "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, **97**, 145-177.

[20] Ma Lingjie, R. Koenker, 2009, "Quantile Regression Methods for Recursive Structural Equation Models," *Journal of Econometrics*, **134**, 471-506.

[21] Manski, C. F. 1988, "Identification of Binary Response Models Source," *Journal of the American Statistical Association* **83:403**, 729-38.

[22] Pollard, D., 1984, *Convergence of Stochastic Processes*, (New York: Springer-Verlag).

[23] Powell, J. L, J. H. Stock and I. M. Stoker, 1989, "Semiparametric Estimation of Index Coefficients," *Econometrica,* **57**, 1403-1430.

[24] Vytlacil, E. and N. Yildiz, 2007, "Dummy Endogenous Models in Weakly Separable Models," *Econometrica* , **75**, 757-779.

[25] Vytlacil, E. and N. Yıldız, 2004, "A Matching Based Estimator of Average Treatment Effect in Weakly Separable Models," manuscript, Yale University and University of Rochester.

A Asymptotic behavior of $\hat{\delta}^{(F)}$:

In this appendix, we study the asymptotic behavior of $\hat{\delta}^{(F)}$ and prove theorem 5.2. For notational simplicity, we will omit the superscript (F) from both $\hat{\beta}$ and $\hat{\delta}$ since these are easily understood to be feasible estimators of the corresponding population parameters. Also, throughout the Appendix we will use $D\rho$ to denote the gradient of the function $\rho : \mathbb{R}^m \to \mathbb{R}$. In addition, throughout the Appendix ϵ_f will be a strictly positive number smaller than of equal to ϵ_f^*.

$$\sqrt{n}\big(\hat{\delta}-\delta\big) = \frac{\frac{1}{\sqrt{n}(n-1)}\sum_i\sum_{j\neq i}\frac{[(X_j-X_i)'\hat{\beta}-\delta]\hat{I}_i\hat{I}_j}{h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\hat{\gamma},\hat{g}_1(Z_i'\hat{\gamma},X_i'\hat{\beta})+\hat{g}_0(Z_j'\hat{\gamma},X_j'\hat{\beta})}{h_{3n}}\right)}{\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}\frac{1}{h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\hat{\gamma},\hat{g}_1(Z_i'\hat{\gamma},X_i'\hat{\beta})+\hat{g}_0(Z_j'\hat{\gamma},X_j'\hat{\beta})}{h_{3n}}\right)\hat{I}_i\hat{I}_j}$$

A.1 The Numerator:

First consider the numerator of this expression:

$$\frac{1}{\sqrt{n}(n-1)}\sum_i\sum_{j\neq i}\frac{1}{h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\hat{\gamma},\hat{g}_1(Z_i'\hat{\gamma},X_i'\hat{\beta})+\hat{g}_0(Z_j'\hat{\gamma},X_j'\hat{\beta})}{h_{3n}}\right)[(X_j-X_i)'\hat{\beta}-\delta]\hat{I}_i\hat{I}_j$$

$$= \frac{1}{\sqrt{n}(n-1)}\sum_i\sum_{j\neq i}\frac{1}{h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)(X_j-X_i)'(\hat{\beta}-\beta)\hat{I}_i\hat{I}_j \tag{21}$$

$$+\frac{1}{\sqrt{n}(n-1)}\sum_i\sum_{j\neq i}\frac{1}{h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)[(X_j-X_i)'\beta-\delta]\hat{I}_i\hat{I}_j \tag{22}$$

$$+\frac{1}{\sqrt{n}(n-1)}\sum_i\sum_{j\neq i}\frac{1}{h_{3n}^2}\left[k\left(\frac{(Z_i-Z_j)'\hat{\gamma},\hat{g}_1(Z_i'\hat{\gamma},X_i'\hat{\beta})+\hat{g}_0(Z_j'\hat{\gamma},X_j'\hat{\beta})}{h_{3n}}\right)\right.$$
$$\left.-k\left(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\right][(X_j-X_i)'\hat{\beta}-\delta]\hat{I}_i\hat{I}_j \tag{23}$$

We will analyze each expression on the right hand side separately.

A.1.1 Analysis of (21):

We first show that

$$\sum_i\sum_{j\neq i}\frac{(X_{jm}-X_{im})}{n(n-1)h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\hat{I}_i\hat{I}_j =$$
$$\sum_i\sum_{j\neq i}\frac{(X_{jm}-X_{im})}{n(n-1)h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)I_iI_j+o_P(1). \tag{24}$$

To do this we first note that

$$\left|\sum_i\sum_{j\neq i}\frac{(X_{jm}-X_{im})}{n(n-1)h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)(\hat{I}_i\hat{I}_j-I_iI_j)\right|\leq$$
$$\sum_i\sum_{j\neq i}\frac{2T_n}{n(n-1)h_{3n}^2}\left|k\left(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)[(\hat{I}_i-I_i]\hat{I}_j+I_i(\hat{I}_j-I_j)]\right| \tag{25}$$

To show that (25) is $o_P(1)$, for $\epsilon_{fn} \leq \epsilon_f^*$ we define

$$\begin{aligned}
\mathcal{H}_n &:= \{f_n : \sup\{|f_n(z'\gamma, x'\beta) - f_{z'\gamma,x'\beta}(z'\gamma, x'\beta)| \leq \epsilon_{fn} : \ w \in [-T_n, T_n]^d\} \\
\mathcal{I}_n &:= \{1\{\tilde{A}\}1\{s \in [-T_n, T_n]^d\} : \tilde{A} = \{(a,b) : f(a,b) \geq q_0\}, f \in \mathcal{H}_n\}.
\end{aligned}$$

Let

$$\begin{aligned}
\overline{A} &:= \{(a,b) : f_{z'\gamma,x'\beta}(a,b) \geq q_0 - \epsilon_{fn}\}, \\
\underline{A} &:= \{(a,b) : f_{z'\gamma,x'\beta}(a,b) \geq q_0 + \epsilon_{fn}\}.
\end{aligned}$$

Next, let $\rho > 0$ and $\eta \in (0,1)$.

$$P\left(\sum_i \sum_{j\neq i} \frac{2T_n}{n(n-1)h_{3n}^2} \left| k\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right) [\hat{I}_j(\hat{I}_i - I_i) + I_i(\hat{I}_j - I_j)] \right| > \rho \right) \leq$$

$$P\left(\sum_i \sum_{j\neq i} \frac{4T_n}{n(n-1)h_{3n}^2} \left| k\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right) (I_i(\overline{A}) - I_i(A)) \right| > \rho, \ \hat{I} \in \mathcal{I}_n \right) + P(\hat{I} \notin \mathcal{I}_n).$$

By equations (17) and (18) there exists N_1 such that the second probability is less than $\frac{\eta}{2}$ for all $n \geq N_1$. Moreover,

$$\begin{aligned}
&P\left(\sum_i \sum_{j\neq i} \frac{4T_n}{n(n-1)h_{3n}^2} \left| k\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right) (I_i(\overline{A}) - I_i(A)) \right| > \rho, \ \hat{I} \in \mathcal{I}_n \right) \\
&\quad < \frac{1}{\rho} E\left(\frac{2T_n}{h_{3n}^2} \left| k\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right) (I_i(\overline{A}) - I_i(A)) \right| \right) \\
&\quad = \frac{2T_n}{\rho} E\left(\int |k(u)| f_{Z\gamma,-g_0}(Z_i\gamma - u_z h_{3n}, g_{1i} - u_g h_{3n}) \mathrm{d}u_z \mathrm{d}u_g |I_i(\overline{A}) - I_i(A)| \right).
\end{aligned}$$

Using continuous differentiability of $f_{Z\gamma,X\beta}$ and $f_{U,\epsilon}$ and compactness of $\overline{A}^*$ the last expression is bounded by

$$c_1 \frac{2T_n h_{3n}}{\rho} + c_2 \frac{2T_n}{\rho} \int\int 1\{q_0 > f_{Z\gamma,X\beta}(Z_i'\gamma, X_i'\beta) \geq q_0 - \epsilon_{fn}\} \mathrm{d}(Z_i'\gamma) \mathrm{d}(X_i'\beta). \tag{26}$$

Since $T_n h_{3n} \to 0$, there exists N_2 such that for all $n \geq N_2$ $c_1 \frac{2T_n h_{3n}}{\rho} \leq \frac{\eta}{4}$. By assumption q_0 has an open neighborhood O such that $\left\|\left(\frac{\partial f_{z'\gamma,x'\beta}(a)}{\partial s_z}, \frac{\partial f_{z'\gamma,x'\beta}(a)}{\partial s_x}\right)\right\| \geq \underline{\theta}$ for each $a \in f^{-1}_{z'\gamma,x'\beta}(O)$. This means that for each such a, at least one of $\left|\frac{\partial f_{z'\gamma,x'\beta}(a)}{\partial(z'\gamma)}\right|$, $\left|\frac{\partial f_{z'\gamma,x'\beta}(a)}{\partial(x'\beta)}\right|$ is greater than or equal to $\underline{\theta}/\sqrt{2}$. If ϵ_f is sufficiently small $[q_0 - \epsilon_{fn}, q_0 + \epsilon_{fn}]$ will be contained in O. Consider the part of $f^{-1}_{z'\gamma,x'\beta}(O)$ where $\left|\frac{\partial f_{z'\gamma,x'\beta}}{\partial(x'\beta)}\right| \geq \frac{\theta}{\sqrt{2}}$. Let B denote this set. Consider the mapping $\varphi : (z'\gamma, f_{z'\gamma,x'\beta}(z'\gamma, x'\beta + \delta)) \to (z'\gamma, x'\beta)$. Let $D := \varphi^{-1}(B)$ and $D_1 := \{z'\gamma : (z'\gamma, x'\beta) \in D\}$. Then

$$\int_{D_1} \int_{q_0 - \epsilon_{fn}}^{q_0 + \epsilon_{fn}} \frac{1}{\left|\frac{\partial f_{z'\gamma,x'\beta}}{\partial(x'\beta)}(s, \varphi_2(s,f))\right|} \mathrm{d}s\mathrm{d}f \leq \frac{2\sqrt{2}\epsilon_{fn}}{\underline{\theta}}.$$

By going through a very similar analysis for the subset of $f^{-1}_{z'\gamma,x'\beta}(O)$, where $\left|\frac{\partial f_{z'\gamma,x'\beta}}{\partial(x'\beta)}\right| \geq \frac{\theta}{\sqrt{2}}$, we would get the same result for that set. This means that the second term in (26) is of the form $C\frac{T_n \epsilon_{fn}}{\rho}$ for some constant C.

Since $\frac{T_n}{\sqrt{n}\tilde{h}_n^3} \to 0$ and $\frac{1}{\tilde{h}_n^2 \log(\tilde{h}_n^{-1})} \to \infty$, $\frac{T_n\sqrt{\log\left(\frac{1}{\tilde{h}_n}\right)}}{\sqrt{n}\tilde{h}_n} \to 0$. Since ϵ_{fn} is $O\left(\sqrt{\frac{1}{n\tilde{h}_n^2}\log\left(\frac{1}{\tilde{h}_n}\right)}\right)$, this means that there exists N_3 such that for all $n \geq N_3$ the last expression in (26) is less than or equal to $\frac{\eta}{4}$. Thus, for all $n \geq \max\{N_1, N_2, N_3\}$ probability that (25) is larger than ρ is less than or equal to η.

Finally, using Lemma 3.1 of PSS and change of variables three times and defining $\rho_m^x(a,b) = E(X_m|Z'\gamma = a, X'\beta = b)$ we can show that

$$\sum_i \sum_{j\neq i} \frac{(X_{jm}-X_{im})}{n(n-1)h_{3n}^2} k\left(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right) I_i I_j =$$
$$E\left[\frac{(X_{jm}-X_{im})}{h_{3n}^2} k\left(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right) I_i I_j\right] + o_P(1) =$$
$$E\left[(\rho_m^x(Z'\gamma, X'\beta+\delta) - \rho_m^x(Z'\gamma, X'\beta))1\{f_{z'\gamma,x'\beta}(Z'\gamma, X'\beta+\delta) \geq q_0\} f_{z'\gamma,x'\beta}(Z'\gamma, X'\beta+\delta)\right] + o_P(1).$$

A.1.2 Analysis of (22):

In this section we show that

$$\sum_i \sum_{j\neq i} \frac{[(X_j-X_i)'\beta-\delta]}{\sqrt{n}(n-1)h_{3n}^2} k\left(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right)(\hat{I}_i\hat{I}_j - I_iI_j) = o_P(1). \tag{27}$$

To do this let $\rho > 0$ and $\eta \in (0,1)$ and choose N_1 sufficiently large so that

$$P\left(\left|\sum_i \sum_{j\neq i} \frac{[(X_j-X_i)'\beta-\delta]}{\sqrt{n}(n-1)h_{3n}^2} k\left(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right)(\hat{I}_i\hat{I}_j - I_iI_j)\right| > \rho\right)$$
$$\leq P\left(\sum_i \sum_{j\neq i}\left|\frac{[(X_j-X_i)'\beta-\delta]}{\sqrt{n}(n-1)h_{3n}^2} k\left(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right)(\hat{I}_i\hat{I}_j - I_iI_j)\right| > \rho,\ \hat{I} \in \mathcal{I}_n\right) + \frac{\eta}{2}$$
$$< \frac{2\sqrt{n}}{\rho} E\left(\left|\frac{[(X_j-X_i)'\beta-\delta]}{h_{3n}^2} k\left(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right) I_i(\overline{A})(I_j(\overline{A}) - I_j(\underline{A}))\right|\right) + \frac{\eta}{2}$$
$$= \frac{\sqrt{n}}{\rho}\int\int\left|\frac{\tilde{g}^{-1}(t,s) - \tilde{g}^{-1}(r,v)}{h_{3n}^2} k\left(\frac{r-t, v+s}{h_{3n}}\right) 1\{f_{z'\gamma,x'\beta}(r, \tilde{g}^{-1}(r,v)-\delta) \geq q_0 - \epsilon_{fn}\}\right.$$
$$\left. f_{z'\gamma,-g_0}(t,s) 1\{q_0 + \epsilon_{fn} \geq f_{z'\gamma,x'\beta}(t, \tilde{g}^{-1}(t,s)) \geq q_0 - \epsilon_{fn}\} f_{z'\gamma,g_1}(r,v) \mathrm{d}t\mathrm{d}s\mathrm{d}r\mathrm{d}v\right|. + \frac{\eta}{2}$$

Using a change of variables, compactness of $\overline{A}^*$ and the support of k and the continuous differentiability of $f_{z'\gamma,x'\beta}$ and $f_{U,\varepsilon}$, and a Taylor expansion we can show that the last expression is bounded by

$$[O(\sqrt{n}h_{3n}) + O(\sqrt{n}h_{3n}^2)]\int\int 1\{q_0 > f_{z'\gamma,x'\beta}(z'\gamma, x'\beta+\delta) \geq q_0 - \epsilon_f\}\mathrm{d}r\mathrm{d}s$$

Using the same arguments as we used at the end of the previous section, we can argue that (27) is $o_P(1)$.

Next, we analyze

$$\frac{1}{\sqrt{n}(n-1)}\sum_i \sum_{j\neq i} \frac{[(X_j-X_i)'\beta-\delta]I_iI_j}{h_{3n}^2} k\left(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}}\right). \tag{28}$$

Adding and subtracting the expectation of each term in the summation and get

$$
\begin{aligned}
&\frac{1}{\sqrt{n}(n-1)}\sum_i\sum_{j\neq i}\Bigg\{\frac{[(X_j-X_i)'\beta-\delta]I_iI_j}{h_{3n}^2}k\bigg(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\bigg)\\
&\qquad -E\left[\frac{[(X_j-X_i)'\beta-\delta]I_iI_j}{h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\right]\Bigg\} \qquad (29)\\
&\qquad +\sqrt{n}E\left[\frac{[(X_j-X_i)'\beta-\delta]I_iI_j}{h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\right]. \qquad (30)
\end{aligned}
$$

We can apply Lemma (3.1) of PSS to the first of these terms. This lemma requires that

$$
\frac{1}{nh_{3n}^2}E\bigg[\frac{1}{h_{3n}^2}k^2\left(\frac{Z_i'\gamma-Z_j'\gamma, g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta}{h_{3n}}\right)[(X_j-X_i)'\beta-\delta]^2I_iI_j\bigg]\to 0.
$$

By arguments similar to those given above, the expectation of the expression in the square brackets is finite. Thus, the condition of the lemma holds if $nh_{3n}^2\to\infty$, and (29) is asymptotically equivalent to

$$
\begin{aligned}
&\tfrac{2}{\sqrt{n}}\textstyle\sum_i\Bigg\{E\left[\tfrac{[(X_j-X_i)'\beta-\delta]I_iI_j}{2h_{3n}^2}k\left(\tfrac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\Bigg|Z_i'\gamma,X_i'\beta\right]\\
&\qquad +E\left[\tfrac{[(X_i-X_j)'\beta-\delta]I_iI_j}{2h_{3n}^2}k\left(\tfrac{(Z_j-Z_i)'\gamma, g_1(Z_j'\gamma,X_j'\beta)+g_0(Z_i'\gamma,X_i'\beta)}{h_{3n}}\right)\Bigg|Z_i'\gamma,X_i'\beta\right]\\
&\qquad -E\left[\tfrac{[(X_j-X_i)'\beta-\delta]I_iI_j}{h_{3n}^2}k\left(\tfrac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\right]\Bigg\}.
\end{aligned}
$$

On the other hand, we have assumed that the kernel function, k, is chosen so that its moments of order one through four are 0, and its fifth absolute moment is finite, that $\tilde{g}^{-1}$ is at least five times continuously differentiable, the density $f_{z'\gamma,x'\beta}$ is four times continuously differentiable, and that $\sqrt{n}h_{3n}^5\to 0$. Under these conditions (30) will approach to 0 as $n\to\infty$. Thus, (28) equals

$$
\begin{aligned}
&\tfrac{1}{\sqrt{n}}\textstyle\sum_i\Bigg\{E\left[\tfrac{[(X_j-X_i)'\beta-\delta]I_iI_j}{h_{3n}^2}k\left(\tfrac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\Bigg|Z_i'\gamma,X_i'\beta\right]\\
&\qquad +E\left[\tfrac{[(X_i-X_j)'\beta-\delta]I_iI_j}{h_{3n}^2}k\left(\tfrac{(Z_j-Z_i)'\gamma, g_1(Z_j'\gamma,X_j'\beta)+g_0(Z_i'\gamma,X_i'\beta)}{h_{3n}}\right)\Bigg|Z_i'\gamma,X_i'\beta\right]\Bigg\}+o_P(1) \qquad (31)
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&E\left(\tfrac{1}{\sqrt{n}}\textstyle\sum_i E\left[\tfrac{[(X_j-X_i)'\beta-\delta]I_iI_j}{h_{3n}^2}k\left(\tfrac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\Bigg|Z_i'\gamma,X_i'\beta\right]\right)^2\\
&\quad = E\Bigg\{E\left[\tfrac{[(X_j-X_i)'\beta-\delta]I_iI_j}{h_{3n}^2}k\left(\tfrac{(Z_i-Z_j)'\gamma, g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\Bigg|Z_i'\gamma,X_i'\beta\right]\Bigg\}^2\\
&= E\bigg[I_i\textstyle\int[\tilde{g}^{-1}(Z_i'\gamma-u_zh_{3n},g_{1i}-u_gh_{3n})-\tilde{g}^{-1}(Z_i'\gamma,g_{1i})]1\{f_{z'\gamma,x'\beta}(Z_i'\gamma-u_zh_{3n},\tilde{g}^{-1}(Z_i'\gamma-u_zh_{3n},g_{1i}-u_gh_{3n}))\geq q_0\}\\
&\qquad\qquad \cdot k(u)f_{z'\gamma,-g_0}(Z_i'\gamma-u_zh_{3n},g_{1i}-u_gh_{3n})\mathrm{d}u_z\mathrm{d}u_g\bigg]^2\to 0,
\end{aligned}
$$

where we used continuity of $\tilde{g}^{-1}$, compactness of the support of k and $\overline{A}$ and the Dominated Convergence Theorem in getting the last result. Similar arguments would show that the $\mathcal{L}^2$ norm of the second sum in (31) converges to 0. Then by Cauchy-Schwarz inequality (31) is $o_P(1)$.

A.1.3 Analysis of (23):

Recall for $r = 0, 1$, we have

$$\sup_{(z'\gamma, x'\beta)} a_n[\hat{g}_r(z'\gamma, x'\beta) - g_r(z'\gamma, x'\beta)]\overline{I}(z'\gamma, x'\beta) \xrightarrow{P} 0. \quad (32)$$

Then as long as $\frac{T_n}{a_n h_{3n}^3} = O(1)$ using arguments as in the analysis of (21) we can argue that (23) equals an $o_P(1)$ term plus

$$\sum_{j\neq i} \frac{[(X_j - X_i)'\beta - \delta]\hat{I}_i\hat{I}_j}{\sqrt{n}(n-1)h_{3n}^2} \left[k\left(\frac{(Z_i - Z_j)'\hat{\gamma}, \hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})}{h_{3n}} \right) - k\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right) \right]. \quad (33)$$

Moreover, by Assumptions (5.4) and (5.7) using second order Taylor expansion yields (23) as

$$\sum_{j\neq i} \frac{[(X_j - X_i)'\beta - \delta]\hat{I}_i\hat{I}_j}{\sqrt{n}(n-1)h_{3n}^2} \left[k\left(\frac{(Z_i - Z_j)'\hat{\gamma}, \hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})}{h_{3n}} \right) - k\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right) \right]$$

$$= \sum_{l=1}^{d_z} \sqrt{n}(\hat{\gamma}_l - \gamma_l) \sum_{j\neq i} \frac{(Z_{il} - Z_{jl})[(X_j - X_i)'\beta - \delta]\hat{I}_i\hat{I}_j}{n(n-1)h_{3n}^3} k_1\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right) \quad (34)$$

$$+ \sum_{j\neq i} \frac{[\hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) - g_1(Z_i'\gamma, X_i'\beta)]\hat{I}_i\hat{I}_j}{\sqrt{n}(n-1)h_{3n}^3} k_2\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right)[(X_j - X_i)'\beta - \delta] \quad (35)$$

$$+ \sum_{j\neq i} \frac{[\hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta}) - g_0(Z_j'\gamma, X_j'\beta)]\hat{I}_i\hat{I}_j}{\sqrt{n}(n-1)h_{3n}^3} k_2\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right)[(X_j - X_i)'\beta - \delta] + o_P(1) \quad (36)$$

where for $l = 1, 2$ k_l denotes l^{th} partial derivative of k.

The analysis of (34) is very similar to the analysis of (21); this term has an extra h_{3n} in the denominator, but it also has the additional term $[(X_j - X_i)'\beta - \delta]$ which is close to zero when g_1 and g_0 are close to one another, and therefore absorbs the extra h_{3n} in the denominator as $n \to \infty$. As a result, (34) equals

$$\sum_{l=1}^{d_z} \sqrt{n}\,(\hat{\gamma}_l - \gamma_l) \sum_{j\neq i} \frac{(Z_{il} - Z_{jl})[(X_j - X_i)'\beta - \delta]I_i I_j}{n(n-1)h_{3n}^3} k_1\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right) + o_P(1).$$

Since $nh_{3n}^4 \to \infty$, Lemma 3.1 of PSS implies that (34) equals

$$\sum_{l=1}^{d_z} \sqrt{n}\,(\hat{\gamma}_l - \gamma_l)\, E\left[\frac{(Z_{il} - Z_{jl})[(X_j - X_i)'\beta - \delta]I_i I_j}{h_{3n}^3} k_1\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right) \right] + o_P(1).$$

Next, we provide the analysis of (35) in detail, but the analysis of (36) will be very similar.

$$\sum_{j\neq i} \frac{\hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) - g_1(Z_i'\gamma, X_i'\beta)}{\sqrt{n}(n-1)h_{3n}^3} k_2\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right)[(X_j - X_i)'\beta - \delta]\hat{I}_i\hat{I}_j$$

$$= \sum_{j\neq i} \frac{\hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) - \hat{g}_1(Z_i'\gamma, X_i'\beta)}{\sqrt{n}(n-1)h_{3n}^3} k_2\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right)[(X_j - X_i)'\beta - \delta]\hat{I}_i\hat{I}_j \quad (37)$$

$$+ \sum_{j\neq i} \frac{\hat{g}_1(Z_i'\gamma, X_i'\beta) - g_1(Z_i'\gamma, X_i'\beta)}{\sqrt{n}(n-1)h_{3n}^3} k_2\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right)[(X_j - X_i)'\beta - \delta]\hat{I}_i\hat{I}_j \quad (38)$$

Since $\frac{T_n^2}{\sqrt{n}h_{3n}^4} \to 0$[14] and $T_n h_{3n} \to 0$, by Taylor's Theorem the first of these expressions equals

$$\sum_{k=1}^{d_z} \sqrt{n}[\hat{\gamma}_k - \gamma_k] \quad \sum_{i_r} \frac{\partial \hat{g}_1(Z_i'\gamma, X_i'\beta)}{\partial(z'\gamma)} \frac{Z_{ik}[(X_j - X_i)'\beta - \delta]\hat{I}_i\hat{I}_j}{n(n-1)h_{3n}^3} k_2\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right)$$
$$+ \sum_{k=1}^{d_z} \sqrt{n}[\hat{\beta}_m - \beta_m] \sum_{i_r} \frac{\partial \hat{g}_1(Z_i'\gamma, X_i'\beta)}{\partial(x'\beta)} \frac{X_{im}[(X_j - X_i)'\beta - \delta]\hat{I}_i\hat{I}_j}{n(n-1)h_{3n}^3} k_2\left(\frac{(Z_i - Z_j)'\gamma, g_1(Z_i'\gamma, X_i'\beta) + g_0(Z_j'\gamma, X_j'\beta)}{h_{3n}} \right) + o_P(1).$$

[14] This follows from $\frac{\sqrt{n}}{a_n^2 h_{3n}^4} \to c < \infty$ and $\frac{T_n a_n}{\sqrt{n}} \to 0$.

By replacing e_1 with e_2 or e_3 in the proof of Theorem 4 of Heckman, Ichimura and Todd (1998) we can show that $\frac{\partial \hat{g}_r(s_z,s_x)}{\partial s_z}$ $(\frac{\partial \hat{g}_r(s_z,s_x)}{\partial s_x})$ is uniformly consistent for $\frac{\partial g_r(s_z,s_x)}{\partial s_z}$ $(\frac{\partial g_r(s_z,s_x)}{\partial s_x})$. Let $\mathcal{G}^1_{1n} := \{\rho \in C^1(\mathbb{R}^2) : \sup_{(z'\gamma,x'\beta)\in\overline{A}} \left|\rho(z'\gamma,x'\beta) - \frac{\partial g_1(z'\gamma,x'\beta)}{\partial(z'\gamma)}\right| \leq \epsilon_{1gn}\}$ with $\epsilon_{1gn} \to 0$. Let $\varpi > 0$, $\eta \in (0,1)$. Then

$$P\left(\frac{\partial \hat{g}_1}{\partial(z'\gamma)} \notin \mathcal{G}^1_{1n},\ \hat{I} \notin \mathcal{I}_n\right) \to 0.$$

In addition,

$$\begin{aligned}
& P\left(\left|\sum_{i_r}\left[\tfrac{\partial \hat{g}_1(Z_i'\gamma,X_i'\beta)}{\partial(z'\gamma)} - \tfrac{\partial g_1(Z_i'\gamma,X_i'\beta)}{\partial(z'\gamma)}\right]\tfrac{Z_{ik}[(X_j-X_i)'\beta-\delta]\hat{I}_i\hat{I}_j}{n(n-1)h^3_{3n}}k_2\left(\tfrac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\right|>\varpi,\ \tfrac{\partial \hat{g}_1}{\partial(z'\gamma)}\in\mathcal{G}^1_{1n},\ \hat{I}\in\mathcal{I}_n\right)\\
\leq\ & P\left(\sum_{i_r}\left|\epsilon_{1gn}\tfrac{T_n[(X_j-X_i)'\beta-\delta]I_i(\overline{A})I_j(\overline{A})}{n(n-1)h^3_{3n}}k_2\left(\tfrac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\right|>\varpi\right)\\
\leq\ & \tfrac{1}{\varpi}E\left(\left|\epsilon_{1gn}\tfrac{T_n[(X_j-X_i)'\beta-\delta]I_i(\overline{A})I_j(\overline{A})}{h^3_{3n}}k_2\left(\tfrac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\right|\right)\to 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}
& P\left(\left|\sum_{i_r}\tfrac{\partial g_1(Z_i'\gamma,X_i'\beta)}{\partial(z'\gamma)}\tfrac{Z_{ik}[(X_j-X_i)'\beta-\delta](\hat{I}_i\hat{I}_j-I_iI_j)}{n(n-1)h^3_{3n}}k_2\left(\tfrac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\right|>\varpi,\ \tfrac{\partial \hat{g}_1}{\partial(z'\gamma)}\in\mathcal{G}^1_{1n},\ \hat{I}\in\mathcal{I}_n\right)\\
\leq\ & \tfrac{1}{\varpi}E\left(\left|\tfrac{\partial g_1(Z_i'\gamma,X_i'\beta)}{\partial(z'\gamma)}\tfrac{T_n[(X_j-X_i)'\beta-\delta]I_i(\overline{A})[I_j(\overline{A})-I_j]}{h^3_{3n}}k_2\left(\tfrac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\right|\right)\to 0.
\end{aligned}$$

These arguments show that

$$\begin{aligned}
&\sum_i\sum_{j\neq i}\tfrac{\partial \hat{g}_1(Z_i'\gamma,X_i'\beta)}{\partial(z'\gamma)}\tfrac{Z_{ik}[(X_j-X_i)'\beta-\delta]\hat{I}_i\hat{I}_j}{n(n-1)h^3_{3n}}k_2\left(\tfrac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right) =\\
&\sum_i\sum_{j\neq i}\tfrac{\partial g_1(Z_i'\gamma,X_i'\beta)}{\partial(z'\gamma)}\tfrac{Z_{ik}[(X_j-X_i)'\beta-\delta]I_iI_j}{n(n-1)h^3_{3n}}k_2\left(\tfrac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right) + o_P(1),
\end{aligned}$$

and that

$$\begin{aligned}
&\sum_i\sum_{j\neq i}\tfrac{\partial \hat{g}_1(Z_i'\gamma,X_i'\beta)}{\partial(x'\beta)}\tfrac{X_{im}[(X_j-X_i)'\beta-\delta]\hat{I}_i\hat{I}_j}{n(n-1)h^3_{3n}}k_2\left(\tfrac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right) =\\
&\sum_i\sum_{j\neq i}\tfrac{\partial g_1(Z_i'\gamma,X_i'\beta)}{\partial(x'\beta)}\tfrac{X_{im}[(X_j-X_i)'\beta-\delta]I_iI_j}{n(n-1)h^3_{3n}}k_2\left(\tfrac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right) + o_P(1).
\end{aligned}$$

Combining these arguments and using Lemma 3.1 of PSS once more, we can show that (37) equals

$$\begin{aligned}
&\sum_{k=1}^{d_z}\sqrt{n}(\hat{\gamma}_k-\gamma_k)E\left[\frac{\partial g_1(Z_i'\gamma,X_i'\beta)}{\partial(z'\gamma)}\frac{Z_{ik}[(X_j-X_i)'\beta-\delta)]I_iI_j}{n(n-1)h^3_{3n}}k_2\left(\frac{(Z_i-Z_j)'\gamma,g_{1i}+g_{0j}}{h_{3n}}\right)\right]\\
+&\sum_{m=1}^{d_x}\sqrt{n}(\hat{\beta}_m-\beta_m)E\left[\frac{\partial g_1(Z_i'\gamma,X_i'\beta)}{\partial(x'\beta)}\frac{X_{im}[(X_j-X_i)'\beta-\delta)]I_iI_j}{n(n-1)h^3_{3n}}k_2\left(\frac{(Z_i-Z_j)'\gamma,g_{1i}+g_{0j}}{h_{3n}}\right)\right] + o_P(1).
\end{aligned}$$

Next we analyze (38). As in the analysis of (22) in Appendix (A.1.2) we could show that this last expression is asymptotically equivalent to

$$\sum_i\sum_{j\neq i}\tfrac{[\hat{g}_1(Z_i'\gamma,X_i'\beta)-g_1(Z_i'\gamma,X_i'\beta)]I_iI_j}{\sqrt{n}(n-1)h^3_{3n}}k_2\left(\tfrac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)[(X_j-X_i)'\beta-\delta].$$

We will analyze this expression using results given in Heckman, Ichimura and Todd (1998) and in Vytlacil and Yildiz (2005). For this purpose, define

$$\hat{\lambda}_n(s_z,s_x) := \sum_{l=1}^{n-1}\frac{[X_l'\beta-s_x-\delta]I_l}{(n-1)h^3_{3n}}k_2\left(\frac{s_z-Z_l'\gamma,g_1(s_z,s_x)+g_0(Z_l'\gamma,X_l'\beta)}{h_{3n}}\right) \tag{39}$$

In the supplementary appendix we prove the following result:

$$\sup_{(s_z,s_x)\in\overline{A}}\left|\hat{\lambda}_n(s_z,s_x)-\lambda_{n0}(s_z.s_x)\right|\overset{a.s.}{\to}0, \tag{40}$$

where

$$\lambda_0(s_z,s_x):=\frac{f_{z'\gamma,x'\beta}(s_z,s_x+\delta)}{f_{U,\varepsilon}(s_z,s_x+\delta)}1\{f_{z'\gamma,x'\beta}(s_z,s_x+\delta)\geq q_0\}. \tag{41}$$

With this result at hand, we could modify the proof of Theorem 3 of Heckman, Ichimura and Todd (1998) to analyze the asymptotic behavior of

$$\textstyle\sum_{i,j\neq i}\frac{[\hat{g}_1(Z_i'\gamma,X_i'\beta)-g_1(Z_i'\gamma,X_i'\beta)]I_iI_j}{\sqrt{n}(n-1)h_{3n}^3}k_2\Big(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\Big)[(X_j-X_i)'\beta-\delta],$$

by rewriting it as

$$\textstyle\sum_i\frac{[\hat{g}_1(Z_i'\gamma,X_i'\beta)-g_1(Z_i'\gamma,X_i'\beta)]I_i}{\sqrt{n}}\hat{\lambda}_n(Z_i'\gamma,X_i'\beta).$$

Now define

$$\Lambda_n:=\{\lambda_n(s_z,s_x):\sup_{(s_z,s_x)\in\overline{A}}|\lambda_n(s_z,s_x)-\lambda_0(s_z,s_x)|\leq\epsilon_l\}.$$

Note that our arguments above show that as n grows large $\hat{\lambda}_n$ will belong to Λ_n with probability approaching to 1, and $\sup_{(s_z,s_x)\in\overline{A}}|\lambda_0(s_z,s_x)|<\infty$. In addition, the assumptions we have made so far imply by Kolmogorov-Tihomirov Lemma that the covering number condition for the Equicontinuity Lemma is satisfied. As a result, a slight modification of the arguments used in the proof of Theorem 3 of Heckman, Ichimura and Todd (1998) yields that

$$\begin{aligned}&\sum_i\sum_{j\neq i}\frac{[\hat{g}_1(Z_i'\gamma,X_i'\beta)-g_1(Z_i'\gamma,X_i'\beta)]I_iI_j}{\sqrt{n}(n-1)h_{3n}^3}k_2\Big(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\Big)[(X_j-X_i)'\beta-\delta]\\&=\sum_i\sum_k\frac{\psi_{ng_1}(D_k,Y_k,Z_k'\gamma,X_k'\beta;Z_i'\gamma,X_i'\beta)}{n^{3/2}}\lambda_0(Z_i'\gamma,X_i'\beta)+\sum_i\frac{\tilde{b}_{g_1}(Z_i'\gamma,X_i'\beta)\lambda_0(Z_i'\gamma,X_i'\beta)}{\sqrt{n}}\\&+\sum_i\frac{\tilde{R}_{g_1}(Z_i'\gamma,X_i'\beta)\lambda_0(Z_i'\gamma,X_i'\beta)}{\sqrt{n}}+o_P(1),\end{aligned}$$

where, writing $m_1(z'\gamma,x'\beta):=E[DY|Z'\gamma=z'\gamma,X'\beta=x'\beta]$, $\epsilon^{g_1}:=DY-m_1(Z'\gamma,X'\beta)$,[15]

$$\psi_{ng_1}(DY,Z'\gamma,X'\beta;z'\gamma,x'\beta)=\frac{e_2[M_{pn}(z'\gamma,x'\beta)]^{-1}I(z'\gamma,x'\beta)\varepsilon^{g_1}}{h_{gn}^3}\left[\left(\frac{(Z'\gamma,X'\beta)-(z'\gamma,x'\beta)}{h_{gn}}\right)^{Q_p}\right]'K^g\left(\frac{(Z'\gamma,X'\beta)-(z'\gamma,x'\beta)}{h_{gn}}\right), \tag{42}$$

$$\begin{aligned}\tilde{b}_{g_1}(Z_i'\gamma,X_i'\beta)=h_{gn}^{\overline{p}-1}e_2[M_p(Z_i'\gamma,X_i'\beta)]^{-1}\textstyle\sum_{k=p+1}^{\overline{p}}\Big[&\textstyle\int u^{Q(0)}\cdot u^{Q(k)}m_1^{(k)}(Z_i'\gamma,X_i'\beta)'\cdot u^{Q(\overline{p}-k)}K_g(u)\mathrm{d}u\\&,\dots,\textstyle\int u^{Q(p)}\cdot u^{Q(k)}m_1^{(k)}(Z_i'\gamma,X_i'\beta)'\cdot u^{Q(\overline{p}-k)}K_g(u)\mathrm{d}u\Big]f_{z'\gamma,x'\beta}^{(\overline{p}-k)}(Z_i'\gamma,X_i'\beta)\hat{I}(Z_i'\gamma,X_i'\beta),\end{aligned} \tag{43}$$

and $n^{-1/2}\sum_{i=1}^n\tilde{R}_{g_1}(Z_i'\gamma,X_i'\beta)\lambda_0(Z_i'\gamma,X_i'\beta)=o_P(1)$. Note that $m_1^{(k)}(s_z,s_x)$, and $f_{z'\gamma,x'\beta}^{(\overline{p}-k)}(s_z,s_x)$ denote the vectors of k^{th} and $(\overline{p}-k)^{th}$ order partial derivatives of m_1 and $f_{z'\gamma,x'\beta}$ evaluated at (s_z,s_x), respectively. Also, M_{pn},

[15]Similarly, $m_0(z'\gamma,x'\beta):=E[(1-D)Y|Z'\gamma=z'\gamma,X'\beta=x'\beta]$, $\epsilon^{g_0}:=(1-D)Y-m_0(Z'\gamma,X'\beta)$.

$u^{Q(s)}$, and $\left(\frac{(Z'\gamma,X'\beta)-(z'\gamma,x'\beta)}{h_{gn}}\right)^{Q_p}$ are as on page 283-284 of Heckman, Ichimura and Todd (1998). In addition, by going through similar arguments as in the analysis of (21) and using the assumption that $\sqrt{n}h_{ng}^{\overline{p}-1} \to c < \infty$ we can show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{b}_{g_1}(Z_i'\gamma,X_i'\beta)\lambda_0(Z_i'\gamma,X_i'\beta)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h_{gn}^{\overline{p}-1}e_2[M_p(Z_i'\gamma,X_i'\beta)]^{-1}\lambda_0(Z_i'\gamma,X_i'\beta)\times$$
$$\sum_{k=p+1}^{\overline{p}}\left[\int u^{Q(0)}\cdot u^{Q(k)}m_1^{(k)}(Z_i'\gamma,X_i'\beta)'\cdot u^{Q(\overline{p}-k)}K_g(u)\mathrm{d}u,\ldots,\int u^{Q(p)}\cdot u^{Q(k)}m_1^{(k)}(Z_i'\gamma,X_i'\beta)'\cdot u^{Q(\overline{p}-k)}K_g(u)\mathrm{d}u\right]f^{(\overline{p}-k)}_{z'\gamma,x'\beta}(Z_i'\gamma,X_i'\beta)I_i+o_P(1)$$

Next, define

$$\tilde{\zeta}_n(D_kY_k,Z_k'\gamma,X_k'\beta;Z_i'\gamma,X_i'\beta):=\frac{e_2[M_{pn}(Z_i'\gamma,X_i'\beta)]^{-1}I_i\varepsilon_k^{g_1}}{h_{gn}^3}\left[\left(\frac{(Z_k'\gamma,X_k'\beta)-(Z_i'\gamma,X_i'\beta)}{h_{gn}}\right)^{Q_p}\right]'$$
$$\times K^g\left(\frac{(Z_k'\gamma,X_k'\beta)-(Z_i'\gamma,X_i'\beta)}{h_{gn}}\right)\lambda_0(Z_i'\gamma,X_i'\beta).$$

and

$$\zeta_n(D_kY_k,Z_k'\gamma,X_k'\beta;D_iY_i,Z_i'\gamma,X_i'\beta)=\tfrac{1}{2}\left[\tilde{\zeta}_n(D_kY_k,Z_k'\gamma,X_k'\beta;Z_i'\gamma,X_i'\beta)+\tilde{\zeta}_n(D_iY_i,Z_i'\gamma,X_i'\beta;Z_k'\gamma,X_k'\beta)\right].$$

Since K^g has compact support, A is compact, and λ_0 is bounded, and $nh_{gn}^6 \to \infty$, all the conditions of Lemma 3.1 of PSS are satisfied and (38) is asymptotically equivalent to

$$\frac{1}{\sqrt{n}}\sum_i E\left[\psi_{ng_1}(D_iY_i,Z_i'\gamma,X_i'\beta;Z_k'\gamma,X_k'\beta)\lambda_0(Z_k'\gamma,X_k'\beta)|Z_i'\gamma,X_i'\beta\right]+b_1.$$

A.2 The denominator:

To analyze the denominator of $\sqrt{n}\left(\hat{\delta}-\delta\right)$, we break it into smaller pieces as well, so that

$$\begin{aligned}
&\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}\frac{1}{h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\hat{\gamma},\hat{g}_1(Z_i'\hat{\gamma},X_i'\hat{\beta})+\hat{g}_0(Z_j'\hat{\gamma},X_j'\hat{\beta})}{h_{3n}}\right)\hat{I}_i\hat{I}_j\\
&=\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}\frac{1}{h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\hat{I}_i\hat{I}_j+ \qquad (44)\\
&\sum_i\sum_{j\neq i}\frac{\hat{I}_i\hat{I}_j}{n(n-1)h_{3n}^2}\left[k\left(\frac{(Z_i-Z_j)'\hat{\gamma},\hat{g}_1(Z_i'\hat{\gamma},X_i'\hat{\beta})+\hat{g}_0(Z_j'\hat{\gamma},X_j'\hat{\beta})}{h_{3n}}\right)\right.\\
&\left.-k\left(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)\right]. \qquad (45)
\end{aligned}$$

Thus, the analysis of the denominator is similar to, but easier than, the analysis of the numerator. Under the conditions we have imposed so far, (45) will vanish with probability approaching to 1, and (44) will be asymptotically equivalent to

$$\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}\frac{1}{h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\gamma,g_1(Z_i'\gamma,X_i'\beta)+g_0(Z_j'\gamma,X_j'\beta)}{h_{3n}}\right)I_iI_j.$$

Furthermore, again, arguments similar to those in the analysis of the numerator can be used to show that this last expression is

$$E[f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)1\{(Z'\gamma,X'\beta)\in\mathcal{S}\cap\mathcal{T}\}],$$

where

$$\begin{aligned}
\mathcal{S} &:= \mathrm{Supp}(Z'\gamma,X'\beta)\cap\mathrm{Supp}(Z'\gamma,X'\beta+\delta),\\
\mathcal{T} &:= \{(s_z,s_x): f_{Z'\gamma,X'\beta}(s_z,s_x)\geq q_0,\ f_{Z'\gamma,X'\beta}(s_z,s_x+\delta)\geq q_0\}.
\end{aligned}$$

A.3 Proof of Theorem 5.2:

Putting the arguments given in sections A.1 and A.2 together and using Change of Variables and Lebesgue Dominated Convergence Theorems and the fact that when k is symmetric $\int\int(-u_g)k_1(u)\mathrm{d}u_z\mathrm{d}u_g = 0 = \int\int(-u_z)k_2(u)\mathrm{d}u_z\mathrm{d}u_g$ and $\int\int(-u_z)k_1(u)\mathrm{d}u_z\mathrm{d}u_g = 1 = \int\int(-u_g)k_2(u)\mathrm{d}u_z\mathrm{d}u_g$. we have

$$\begin{aligned}\sqrt{n}\left(\hat{\delta}^{(F)} - \delta\right) = &\tfrac{1}{E[f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)1\{(Z'\gamma,X'\beta)\in\mathcal{S}\cap\mathcal{T}\}]}\times \\ &\sqrt{n}\left(\hat{\gamma}-\gamma\right)'(\alpha_\gamma+\nu_{1\gamma}+\nu_{0\gamma}) + \sqrt{n}\left(\hat{\beta}-\beta\right)'(\alpha_\beta+\nu_{1\beta}+\nu_{0\beta}) \\ &+\textstyle\sum_i \frac{1}{\sqrt{n}}E\left[\psi_{ng_1}(D_k,Y_k,Z_k'\gamma,X_k'\beta;Z_i'\gamma,X_i'\beta)|Z_i'\gamma,X_i'\beta\right]\lambda_0(Z_i'\gamma,X_i'\beta)+b_1 \\ &+\textstyle\sum_i \frac{1}{\sqrt{n}}E\left[\psi_{ng_0}(D_k,Y_k,Z_k'\gamma,X_k'\beta;Z_i'\gamma,X_i'\beta)|Z_i'\gamma,X_i'\beta\right]\lambda_0(Z_i'\gamma,X_i'\beta)+b_0\Big) + o_P(1),\end{aligned}$$

where for $m = 1,\dots,d_z$, $l = \dots d_x$, $k = 0,1$, $\rho^z_m(a,b) = E(Z_{im}|Z_i'\gamma = a, X_i'\beta = b)$, $\rho^x_l(a,b) = E(X_{il}|Z_i'\gamma = a, X_i'\beta = b)$, and

$$\alpha_{\beta l} = E\Big[(\rho^x_l(Z'\gamma,X'\beta+\delta)-\rho^x_l(Z'\gamma,X'\beta))1\{f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta)\geq q_0\}1\{f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)\geq q_0\}f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)\Big],$$

$$\alpha_{\gamma m} = E\left[\left(\rho^z_m(Z'\gamma,X'\beta)-\rho^z_m(Z'\gamma,X'\beta+\delta)\right)\frac{\left(-\int_{-\infty}^{X'\beta+\delta}\frac{\partial f_{U,\varepsilon}(Z'\gamma,e)\mathrm{d}e}{\partial u}\right)}{f_{U\varepsilon}(Z'\gamma,X'\beta+\delta)}1\{f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta)\geq q_0\}1\{f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)\geq q_0\}f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)\right],$$

$$\nu_{k\gamma m} = E\left[\rho^z_m(Z'\gamma,X'\beta)\frac{\int_{-\infty}^{X'\beta+\delta k}\frac{\partial f_{U,\varepsilon}(Z'\gamma,e)\mathrm{d}e}{\partial u}}{f_{U\varepsilon}(Z'\gamma,X'\beta+\delta)}1\{f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+)\geq q_0\}1\{f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)\geq q_0\}f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)\right],$$

$$\nu_{k\beta l} = E\left[\rho^x_l(Z'\gamma,X'\beta)\frac{f_{U,\varepsilon}(Z'\gamma,X'\beta)}{f_{U\varepsilon}(Z'\gamma,X'\beta+\delta)}(1-k)1\{f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta)\geq q_0\}1\{f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)\geq q_0\}f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)\right].$$

Since $\sqrt{n}(\hat{\gamma}-\gamma) = \frac{1}{\sqrt{n}}\sum_i \psi^\gamma(D_i,Z_i) + o_p(1)$ and $\sqrt{n}(\hat{\beta}-\beta) = \frac{1}{\sqrt{n}}\sum_i \psi^\beta(Y_i,Z_i'\gamma,X_i) + o_p(1)$,

$$\sqrt{n}\left(\hat{\delta}^{(F)}-\delta\right) = \textstyle\sum_i \frac{1}{\sqrt{n}}\psi^\delta(Y_i,D_i,Z_i,X_i) + \frac{b_1+b_0}{E[f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)1\{(Z'\gamma,X'\beta)\in\mathcal{S}\cap\mathcal{T}\}]} + o_P(1),$$

where

$$\begin{aligned}\psi^\delta(Y_i,D_i,Z_i,X_i) := &\frac{[\psi^\gamma(D_i,Z_i)]'(\alpha_\gamma+\nu_{1\gamma}+\nu_{0\gamma})+[\psi^\beta(Y_i,Z_i'\gamma,X_i)]'(\alpha_\beta+\nu_{1\beta}+\nu_{0\beta})}{E[f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)1\{(Z'\gamma,X'\beta)\in\mathcal{S}\cap\mathcal{T}\}]} \\ &+\lambda_0(Z_i'\gamma,X_i'\beta)\frac{E\left[\psi_{ng_1}(D_k,Y_k,Z_k'\gamma,X_k'\beta;Z_i'\gamma,X_i'\beta)|Z_i'\gamma,X_i'\beta\right]+E\left[\psi_{ng_0}(D_k,Y_k,Z_k'\gamma,X_k'\beta;Z_i'\gamma,X_i'\beta)|Z_i'\gamma,X_i'\beta\right]}{E[f_{z'\gamma,x'\beta}(Z'\gamma,X'\beta+\delta)1\{(Z'\gamma,X'\beta)\in\mathcal{S}\cap\mathcal{T}\}]}.\end{aligned}$$

Thus, the variance of the limiting distribution of $\sqrt{n}\left(\hat{\delta}^{(F)}-\delta\right)$ will be $E[(\psi^\delta(Y_i,D_i,Z_i,X_i))^2]$. This variance could be estimated by

$$\frac{1}{n}\sum_i(\hat{\psi}^\delta(Y_i,D_i,Z_i,X_i))^2$$

where

$$\begin{aligned}\hat{\psi}^\delta(Y_i,D_i,Z_i,X_i) := &\frac{[\hat{\psi}^\gamma(D_i,Z_i)]'(\hat{\alpha}_\gamma+\hat{\nu}_{1\gamma}+\hat{\nu}_{0\gamma})+[\hat{\psi}^\beta(Y_i,Z_i'\gamma,X_i)]'(\hat{\alpha}_\beta+\hat{\nu}_{1\beta}+\hat{\nu}_{0\beta})}{\sum_i\sum_j\frac{(X_j-X_i)'\hat{\beta}\hat{I}_i\hat{I}_j}{n(n-1)h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\hat{\gamma},\hat{g}_1(Z_i'\hat{\gamma},X_i'\hat{\beta})+\hat{g}_0(Z_i'\hat{\gamma},X_i'\hat{\beta})}{h_{3n}}\right)} \\ &+\hat{\lambda}(Z_i'\hat{\gamma},X_i'\hat{\beta})\frac{\sum_{k\neq i}\frac{\hat{\psi}_{ng_1}(D_k,Y_k,Z_k'\hat{\gamma},X_k'\hat{\beta};Z_i'\hat{\gamma},X_i'\hat{\beta})+\hat{\psi}_{ng_0}(D_k,Y_k,Z_k'\hat{\gamma},X_k'\hat{\beta};Z_i'\hat{\gamma},X_i'\hat{\beta})}{(n-1)}}{\sum_i\sum_j\frac{(X_j-X_i)'\hat{\beta}\hat{I}_i\hat{I}_j}{n(n-1)h_{3n}^2}k\left(\frac{(Z_i-Z_j)'\hat{\gamma},\hat{g}_1(Z_i'\hat{\gamma},X_i'\hat{\beta})+\hat{g}_0(Z_i'\hat{\gamma},X_i'\hat{\beta})}{h_{3n}}\right)}.\end{aligned}$$

We next discuss what the estimated quantities in the definition of $\hat{\psi}^\delta$ are. Section 5.1 gives estimators for ψ^γ (i.e. $\hat{r}_{n\tilde{\gamma}}(Z_i,D_i)$) and ψ^β (i.e. $\hat{r}_{n\tilde{\beta}}(X_i,Z_i,D_i,Y_i)-2\hat{r}_{n\tilde{\gamma}}(Z_i,D_i)\hat{C}$). Moreover, we can estimate α_γ, α_β, $\nu_{r\gamma}$ and $\nu_{r\beta}$ for

$r = 0, 1$ by

$$
\begin{aligned}
\hat{\alpha}_\gamma &= \sum_i \sum_{j \neq i} \frac{(Z_i - Z_j)[\hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})]\hat{I}_i\hat{I}_j}{n(n-1)h_{3n}^3} k_1\left(\frac{(Z_i - Z_j)'\hat{\gamma}, \hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})}{h_{3n}}\right) \\
\hat{\alpha}_\beta &= \sum_i \sum_{j \neq i} \frac{(X_i - X_j)\hat{I}_i\hat{I}_j}{n(n-1)h_{3n}^2} k\left(\frac{(Z_i - Z_j)'\hat{\gamma}, \hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})}{h_{3n}}\right) \\
\hat{\nu}_{r\gamma} &= \sum_i \sum_{j \neq i} \frac{\partial \hat{g}_r(Z_i'\hat{\gamma}, X_i'\hat{\beta})}{\partial(z'\gamma)} \frac{Z_i[\hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})]\hat{I}_i\hat{I}_j}{n(n-1)h_{3n}^3} k_2\left(\frac{(Z_i - Z_j)'\hat{\gamma}, \hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})}{h_{3n}}\right) \\
\hat{\nu}_\beta &= \sum_i \sum_{j \neq i} \frac{\partial \hat{g}_r(Z_i'\hat{\gamma}, X_i'\hat{\beta})}{\partial(x'\beta)} \frac{X_i[\hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})]\hat{I}_i\hat{I}_j}{n(n-1)h_{3n}^3} k_2\left(\frac{(Z_i - Z_j)'\hat{\gamma}, \hat{g}_1(Z_i'\hat{\gamma}, X_i'\hat{\beta}) + \hat{g}_0(Z_j'\hat{\gamma}, X_j'\hat{\beta})}{h_{3n}}\right).
\end{aligned}
$$

Also λ_0 could be estimated by $\hat{\lambda}$ defined in Equation (39), and for $r = 0, 1$, $\psi_{ng_r}(D_k, Y_k, Z_k'\gamma, X_k'\beta; Z_i'\gamma, X_i'\beta)$ can be estimated by

$$\hat{\psi}_{ng_r}(D,Y,Z'\gamma,X'\beta;Z_i'\hat{\gamma},X_i'\hat{\beta}) = \frac{e_2[\hat{M}_{pn}(Z_i'\hat{\gamma},X_i'\hat{\beta})]^{-1}\hat{I}_i\hat{\epsilon}^{g_r}}{h_{gn}^3}\left[\left(\frac{(Z'\gamma,X'\beta)-(Z_i'\hat{\gamma},X_i'\hat{\beta})}{h_{gn}}\right)^{Q_p}\right]' K^g\left(\frac{(Z'\gamma,X'\beta)-(Z_i'\hat{\gamma},X_i'\hat{\beta})}{h_{gn}}\right),$$

where $\hat{\epsilon}^{g_1} = DY - \hat{E}(DY|Z_i'\hat{\gamma}, X_i'\hat{\beta}]$ and $\hat{\epsilon}^{g_0} = (1-D)Y - \hat{E}((1-D)Y|Z_i'\hat{\gamma}, X_i'\hat{\beta}]$, and $\hat{M}_{pn}$ and $\left(\frac{(Z'\gamma,X'\beta)-(Z_i'\hat{\gamma},X_i'\hat{\beta})}{h_{gn}}\right)^{Q_p}$ are as on page 283, 284 of Heckman, Ichimura and Todd (1998).

B Asymptotic behavior of $\hat{\beta}^{(F)}$:

B.1 Assumptions needed for $\hat{\beta}^{(Inf)}$:

Our Assumptions (C-1) through (C-6) are the same as Assumptions 1-6 of PSS adopted to our setting. [16]

(C-1) The support $\Omega^{(x,z'\tilde{\gamma})}$ of $(X', Z'\tilde{\gamma})'$ is a convex subset of $\mathbb{R}^{d_x+1}$ with nonempty interior $\Omega_0^{(x,z'\tilde{\gamma})}$.

(C-2) The density, $f_{x,z'\tilde{\gamma}}(x,t)$, of $(X', Z'\tilde{\gamma})'$ is continuous in $(x',t)'$ for all $(x',t)' \in \mathbb{R}^{d_x+1}$, so that $f_{x,z'\tilde{\gamma}}(x,t) = 0$ for all $(x',t) \in \partial\Omega^{(x,z'\tilde{\gamma})}$, where $\partial\Omega^{(x,z'\tilde{\gamma})}$ denotes the boundary of $\Omega^{(x,z'\tilde{\gamma})}$. Furthermore, $f_{x,z'\tilde{\gamma}}$ is continuously differentiable for all $(x',t) \in \Omega_0^{(x,z'\tilde{\gamma})}$ and $\psi(x,t)$ is continuously differentiable in $(x',t)'$ for all $(x',t)' \in \overline{\Omega}^{(x,z'\tilde{\gamma})}$, where $\overline{\Omega}^{(x,z'\tilde{\gamma})}$ differs from $\Omega_0^{(x,z'\tilde{\gamma})}$ by a set of measure 0.

(C-3) The components of the random vector $\partial\psi(X, Z'\tilde{\gamma})/\partial x$ and random matrix $[\partial f_{x,z'\tilde{\gamma}}/\partial x][Y, X', Z'\tilde{\gamma}]$ have finite second moments. Also, $\partial f_{x,z'\tilde{\gamma}}/\partial x$ and $\partial(\psi f_{x,z'\tilde{\gamma}})/\partial x$ satisfy the following Lipschitz conditions: For some $m_2(x,t)$,

$$\left|\left|\frac{\partial f_{x,z'\tilde{\gamma}}(x+s_x,t+s_t)}{\partial x} - \frac{\partial f_{x,z'\tilde{\gamma}}(x,t)}{\partial x}\right|\right| < m_2(x,t)||s||,$$

$$\left|\left|\frac{\partial[f_{x,z'\tilde{\gamma}}(x+s_x,t+s_t)\psi(x+s_x,t+s_t)]}{\partial x} - \frac{\partial[f_{x,z'\tilde{\gamma}}(x,t)\psi(x,t)]}{\partial x}\right|\right| < m_2(x,t)||s||,$$

with $E[(1 + Y + ||X, Z'\tilde{\gamma}||)m_2(X, Z'\tilde{\gamma})]^2 < \infty$.

[16] These assumptions hold if and only if the corresponding assumptions for $f_{x,z'\gamma}$ hold.

(C-4) The support $\Omega_K^{x,t}$ of $K_{x,t}(u)$ is a convex subset of $\mathbb{R}^{d_x+1}$ with nonempty interior, with the origin as an interior point. $K_{x,t}(u)$ is a bounded differentiable function such that $\int K_{x,t}(u)\mathrm{d}u = 1$ and $\int uK_{x,t}(u)\mathrm{d}u = 0$. $K_{x,t}(u) = 0$ for all $u \in \partial\Omega_K^{x,t}$, where $\partial\Omega_K^{x,t}$ denotes the boundary of $\Omega_K^{x,t}$. $K_{x,t}(u)$ is a symmetric function; $K_{x,t}(u) = K_{x,t}(-u)$ for all $u \in \Omega_K^{x,t}$.

(C-5) Let $s_{xt} := (d_x+5)/2$ if d_x is odd and $s_{xt} := (d_x+4)/2$ if d_x is even.[17] All partial derivatives of $f_{x,z'\tilde{\gamma}}(x,t)$ of order $s_{xt}+1$ exist. The expectation $E[Y(\partial^\rho f_{x,z'\tilde{\gamma}}(x,t)/\partial z_{l_1}...\partial z_{l_\rho})]$ exists for all $\rho \leq s_{x,t}+1$. All moments of $K_{x,t}(u)$ of order $s_{x,t}$ exist.

(C-6) The kernel function $K_{x,t}(\cdot)$ obeys

$$\begin{array}{lll} \int u_1^{l_1}...u^{l_{\rho'}} K_{x,t}(u)\mathrm{d}u = 0 & \text{for} & l_1 + ... + l_{\rho'} < s_{x,t}, \text{ and} \\ \int u_1^{l_1}...u^{l_{\rho'}} K_{x,t}(u)\mathrm{d}u \neq 0 & \text{for} & l_1 + ... + l_{\rho'} = s_{x,t}. \end{array}$$

B.2 Proof of Lemma (5.2):

To analyze the difference in expectations, we use the Mean Value Theorem to write for some $\tilde{\xi}$ between $\hat{\xi}$ and γ

$$\begin{aligned}
&\Bigg|\tfrac{1}{n(n-1)}\textstyle\sum_i\sum_{j\neq i}\Bigg\{\left[\tfrac{-2}{h_{2n}^{d_x+3}}\tfrac{\partial K_x}{\partial x_r}\left(\tfrac{X_i-X_j}{h_{2n}}\right)K_t'\left(\tfrac{Z_i'\hat{\xi}-Z_j'\hat{\xi}}{h_{2n}}\right)(Z_{il}-Z_{jl})Y_i\right] \\
&\qquad -\left[\tfrac{-2}{h_{2n}^{d_x+3}}\tfrac{\partial K_x}{\partial x_r}\left(\tfrac{X_i-X_j}{h_{2n}}\right)K_t'\left(\tfrac{Z_i'\gamma-Z_j'\gamma}{h_{2n}}\right)(Z_{il}-Z_{jl})Y_i\right]\Bigg\}\Bigg| \\
&= \left|\textstyle\sum_i\sum_{j\neq i}\sum_{m=1}^{d_z}\left[\tfrac{2Y_i(Z_{im}-Z_{jm})(\hat{\xi}_m-\gamma_m)(Z_{il}-Z_{jl})}{n(n-1)h_{2n}^{d_x+4}}\tfrac{\partial K_x}{\partial x_r}\left(\tfrac{X_i-X_j}{h_{2n}}\right)K_t''\left(\tfrac{Z_i'\tilde{\xi}-Z_j'\tilde{\xi}}{h_{2n}}\right)\right]\right| \\
&\leq \textstyle\sum_{m=1}^{d_z}\sqrt{n}|\hat{\gamma}_m-\gamma_m|\sum_i\sum_{j\neq i}\tfrac{2R_{2t}}{\sqrt{n}n(n-1)h_{2n}^{d_x+4}}\left[\left|\tfrac{\partial K_x}{\partial x_r}\left(\tfrac{X_i-X_j}{h_{2n}}\right)\right||Z_{im}-Z_{jm}||Z_{il}-Z_{jl}|\right] \\
&\leq \left(\textstyle\sum_{m=1}^{d_z}\sqrt{n}|\hat{\gamma}_m-\gamma_m|\right)\sum_i\sum_{j\neq i}\tfrac{2R_{2t}}{\sqrt{n}n(n-1)h_{2n}^{d_x+4}}\left[\left|\tfrac{\partial K_x}{\partial x_r}\left(\tfrac{X_i-X_j}{h_{2n}}\right)\right|\|Z_i-Z_j\|\right],
\end{aligned}$$

where R_{2t} denotes the bound on K_t''. To write this last inequality, we used $|Y| \leq 1$, boundedness of K_t'' and that $\hat{\xi}$ lies between $\hat{\gamma}$ and γ. We know that $\sqrt{n}|\hat{\gamma}_m - \gamma_m| = O_P(1)$ for each m. Using change of variables, boundedness of derivatives of K_{xt}, the Lipschitz condition on $E(\|Z_i\|^2|X_i, Z_i'\gamma)$ and $nh_{2n}^8 \to \infty$, we can show that the other term converges to 0 in $\mathcal{L}_1$, and thus, is $o_P(1)$. This completes the proof of Lemma (5.2). ■

B.3 Proof of Theorem (5.1):

The analysis given in the text and the analysis on pages 1411-1412 of PSS applied to our setting tells us that

$$\sqrt{n}\left(\hat{\beta}^{(F)}-\beta\right)' = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\Big\{r_\beta(X_i, Z_i'\gamma) - 2r_\gamma(Z_i)C - \left[Er_\beta(X, Z'\gamma) - 2Er_\gamma(Z)C\right]\Big\} + o_P(1),$$

where C is the $d_z \times d_x$ matrix whose l^{th}-row equals C_l. Applying the Central Limit Theorem to this expression we get that the limiting distribution of $\sqrt{n}\left(\hat{\beta}^{(F)}-\beta\right)$ is normal with mean 0 and variance equals
$4E[r_\beta(X,Z'\gamma)'r_\beta(X,Z'\gamma)] - 16E[r_\beta(X,Z'\gamma)'r_\gamma(Z)]C + 16C'Er_\gamma(Z)'r_\gamma(Z)]C - 4\beta\beta' + 16\beta\gamma'C - 16C'\gamma\gamma'C.$ ■

[17] We would like to make sure that $nh_{2n}^{2s_{x,t}} \to 0$ and that $nh_{2n}^{d_x+3} \to \infty$. These two conditions will hold jointly only if $2s_{x,t} > d_x + 3$. The $s_{x,t}$ given in this assumption is the smallest integer satisfying this last condition.

Theorem B.1 *$\Sigma_{\hat{\tilde{\beta}}^{(F)}}$ can be consistently estimated by $\hat{\Sigma}_{\tilde{\beta}} + 4\hat{C}'\hat{\Sigma}_{\tilde{\gamma}}\hat{C} - 4\hat{\Sigma}_{\tilde{\beta}\tilde{\gamma}}\hat{C}$ with*
$\hat{\Sigma}_{\tilde{\gamma}} = 4\frac{\sum_i \hat{r}_{n\tilde{\gamma}}(Z_i,D_i)'\hat{r}_{n\tilde{\gamma}}(Z_i,D_i)}{n} - 4\hat{\tilde{\gamma}}\hat{\tilde{\gamma}}'$, $\hat{\Sigma}_{\tilde{\beta}} = 4\frac{\sum_i \hat{r}_{n\tilde{\beta}}(X_i,Z_i,D_i,Y_i)'\hat{r}_{n\tilde{\beta}}(X_i,Z_i,D_i,Y_i)}{n} - 4\hat{\tilde{\beta}}^{(F)}\hat{\tilde{\beta}}^{(F)\prime}$,
$\hat{\Sigma}_{\tilde{\beta}\tilde{\gamma}} = 4\frac{\sum_i \hat{r}_{n\tilde{\beta}}(X_i,Z_i,D_i,Y_i)'\hat{r}_{n\tilde{\gamma}}(Z_i,D_i)}{n} - 4\hat{\tilde{\beta}}^{(F)}\hat{\tilde{\gamma}}'$, *$\hat{C}_l$ is as defined in Theorem 5.2,*
$\hat{r}_{n\tilde{\gamma}}(Z_i,D_i) = \frac{-1}{(n-1)h_{\gamma n}^{d_z+1}}\sum_{j\neq i}\frac{\partial K_\gamma}{\partial z}\left(\frac{Z_i-Z_j}{h_{\gamma n}}\right)(D_i-D_j)$, *and*
$\hat{r}_{n\tilde{\beta}}(X_i,Z_i,D_i,Y_i) = \frac{-1}{(n-1)h_{\beta n}^{d_x+2}}\sum_{j\neq i}\frac{\partial K}{\partial t}\left(\frac{(X_i,Z_i'\hat{\tilde{\gamma}})-(X_j,Z_j'\hat{\tilde{\gamma}})}{h_{2n}}\right)(Y_i-Y_j)$

Proof B.1 *The fact that $\hat{\Sigma}_{\hat{\beta}}, \hat{\Sigma}_{\hat{\gamma}}$ are consistent for $\Sigma_\beta, \Sigma_\gamma$ follows from the analysis of PSS. Consistency of $\hat{C}$ for C is shown in the main text.* ■